Modern Time-Series Forecasting as Enabler of Proactive Systems Management

Samuel Kounev

Chair of Software Engineering
University of Würzburg

http://descartes-research.net/

Seminar, on Modern Distributed Systems, Amsterdam, June 12, 2018

Selected References

- S. Kounev, J. O. Kephart, A. Milenkoski, and X. Zhu. (eds.) Self-Aware Computing Systems. Springer Verlag, Berlin Heidelberg, Germany, 2017. http://www.springer.com/de/book/9783319474724
- N. Huber, F. Brosig, S. Spinner, S. Kounev, and M. Bähr. **Model-Based Self-Aware Performance and Resource Management Using the Descartes Modeling Language**. *IEEE Transactions on Software Engineering (TSE)*, PP(99), 2017, IEEE Computer Society. To appear. [pdf | DOI | http]
- S. Kounev, N. Huber, F. Brosig, and X. Zhu. **A Model-Based Approach to Designing Self-Aware IT Systems and Infrastructures**. *IEEE Computer*, 49(7):53–61, July 2016, IEEE. [pdf | DOI | http]
- S. Kounev, F. Brosig, and N. Huber. **The Descartes Modeling Language**. Technical report, Department of Computer Science, University of Wuerzburg, October 2014. [http | http | .pdf]
- F. Brosig, N. Huber, and S. Kounev. **Architecture-Level Software Performance Abstractions for Online Performance Prediction**. *Elsevier Science of Computer Programming Journal (SciCo)*, Vol. 90, Part B:71-92, 2014, Elsevier. [DOI | http | .pdf]
- N. Huber, A. van Hoorn, A. Koziolek, F. Brosig, and S. Kounev. **Modeling Run-Time Adaptation at the System Architecture Level in Dynamic Service-Oriented Environments**. *Service Oriented Computing and Applications Journal (SOCA)*, 8(1):73-89, 2014, Springer-Verlag. [DOI | .pdf]
- F. Brosig, P. Meier, S. Becker, A. Koziolek, H. Koziolek, and S. Kounev. **Quantitative Evaluation of Model-Driven Performance Analysis and Simulation of Component-based Architectures**. *IEEE Transactions on Software Engineering (TSE)*, 41(2):157-175, February 2015, IEEE. [DOI | http | .pdf]
- F. Gorsler, F. Brosig, and S. Kounev. **Performance Queries for Architecture-Level Performance Models**. In *5th ACM/SPEC International Conference on Performance Engineering (ICPE 2014)*, Dublin, Ireland, 2014. ACM, New York, NY, USA. 2014. [DOI | .pdf]
- N. Herbst, N. Huber, S. Kounev and E. Amrehn. **Self-Adaptive Workload Classification and Forecasting for Proactive Resource Provisioning**. *Concurrency and Computation - Practice and Experience, John Wiley and Sons, Ltd.*, 26(12):2053-2078, 2014. [DOI | http | .pdf]
- S. Spinner, G. Casale, F. Brosig, and S. Kounev. **Evaluating Approaches to Resource Demand Estimation**. *Performance Evaluation*, 92:51 - 71, October 2015, Elsevier B.V. [DOI | http | .pdf]
- *N. Herbst, S. Kounev and R. Reussner.* ***Elasticity: What it is, and What it is Not.*** *In 10th Intl. Conference on Autonomic Computing (ICAC 2013), San Jose, CA, June 24-28, 2013.* [slides | http | .pdf]
- A. Milenkoski, M. Vieira, S. Kounev, A. Avrtizer, and B. Payne. **Evaluating Computer Intrusion Detection Systems: A Survey of Common Practices**. *ACM Computing Surveys*, 48(1):12:1-12:41, September 2015, ACM, New York, NY, USA. **5-year Impact Factor (2014): 5.949**. [http]

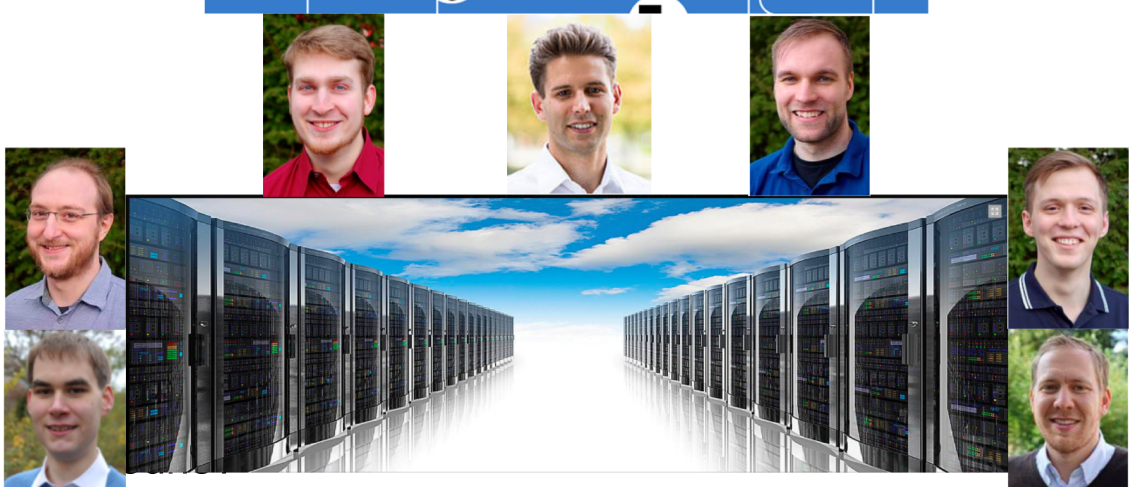
Chair of Software-Engineering (March 2018)

Research Landscape

IT Security

Internet of
Things (IoT)

Cyber-Physical
Systems (CPS)



Cloud
Computing

Inspiration vs. Perspiration

- "Whoever has visions should go to the doctor."

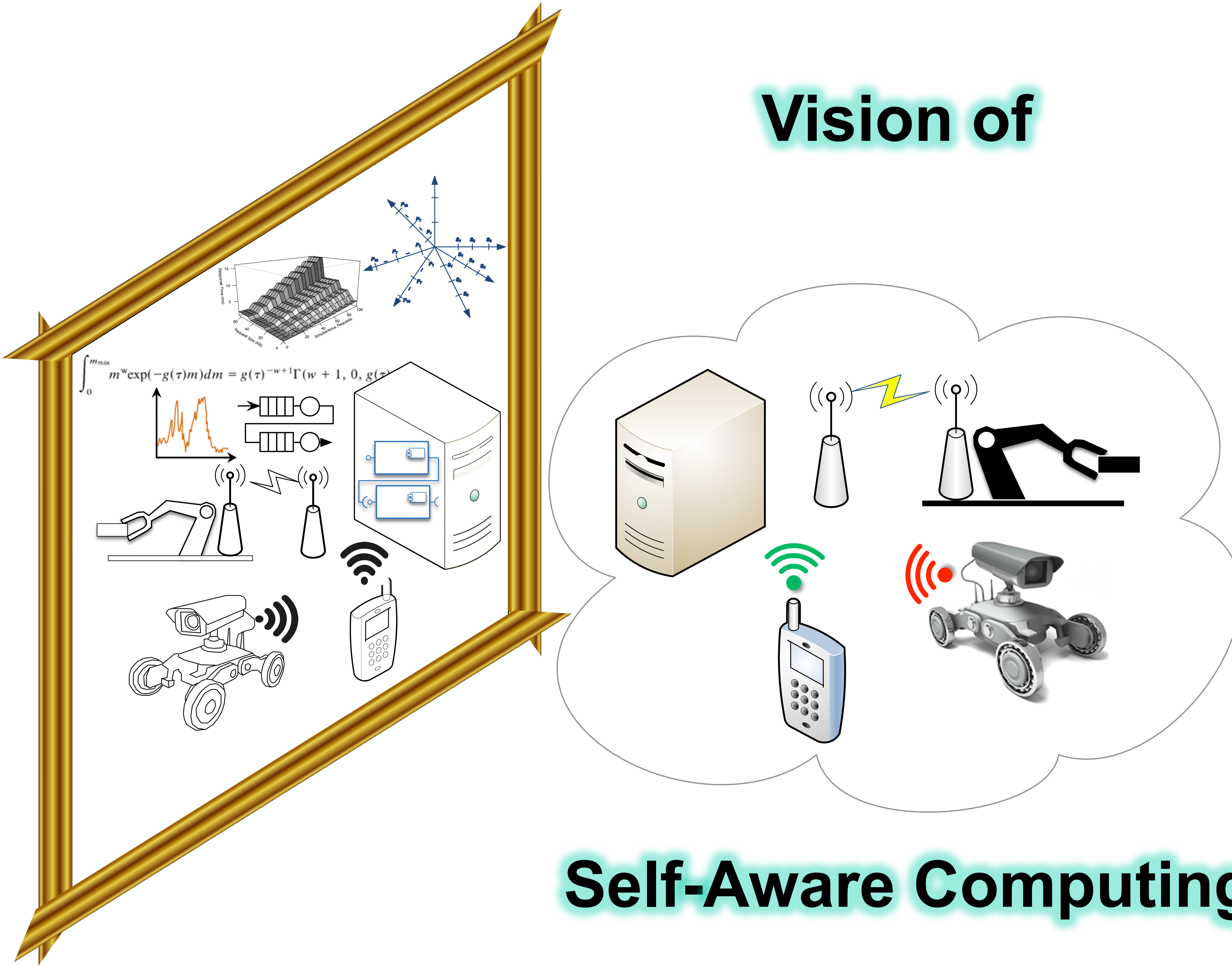
Helmut Schmidt

QuotePixel.com

„Mit Träumen beginnt die Realität."

Christoph Daum (1953*), Fußballspieler und -trainer

Vision of Self-Aware Computing



Dagstuhl-Seminar

Model-driven Algorithms and Architectures for Self-Aware Computing Systems, Jan 18-23, 2015, Dagstuhl Seminar 15041

Organizers

Jeffrey O. Kephart (IBM TJ Watson Research Center, US)
Samuel Kounev (Universität Würzburg, DE)
Marta Kwiatkowska (University of Oxford, GB)
Xiaoyun Zhu (VMware, Inc., US)

Community:
http://descartes.tools/self-aware
Dagstuhl Report:
http://drops.dagstuhl.de/opus/volltexte/2015/5038/
Seminar Page:
http://www.dagstuhl.de/15041

Schloss Dagstuhl
Where Computer Scientists Meet

Book

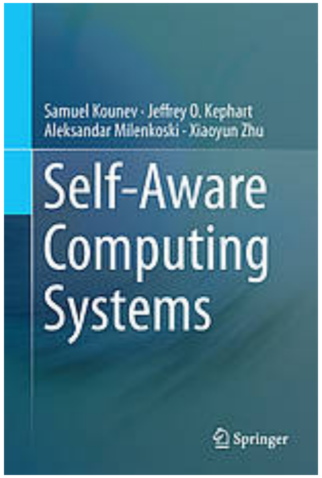
- **„Self-Aware Computing Systems“**

  Samuel Kounev (University of Würzburg, DE)

  Jeffrey O. Kephart (IBM T.J. Watson, USA)

  Aleksandar Milenkoski (University of Würzburg, DE)

  Xiaoyun Zhu (Futurewei Technologies, Huawei, USA)



- 27 chapters, ca 700 pages, ca. 50 authors involved

S. Kounev, J. O. Kephart, A. Milenkoski, and X. Zhu. (eds.) **Self-Aware Computing Systems**. Springer Verlag, Berlin Heidelberg, Germany, 2017. http://www.springer.com/de/book/9783319474724

Definition

Self-Aware Computing Systems are computing systems that:

1. ***learn models*** capturing knowledge about themselves and their environment ***on an ongoing basis*** and

2. ***reason*** using the models enabling them to ***act*** based on their knowledge and reasoning

in accordance with ***higher-level goals***, which may also be subject to change.

S. Kounev, P. Lewis, K. Bellman, N. Bencomo, J. Camara, A. Diaconescu, L. Esterle, K. Geihs, H. Giese, S. Goetz, P. Inverardi, J. Kephart and A. Zisman. **The Notion of Self-Aware Computing**. In Self-Aware Computing Systems, S. Kounev, J. O. Kephart, A. Milenkoski, and X. Zhu, editors. Springer Verlag, Berlin Heidelberg, Germany, 2017.

Extended Definition

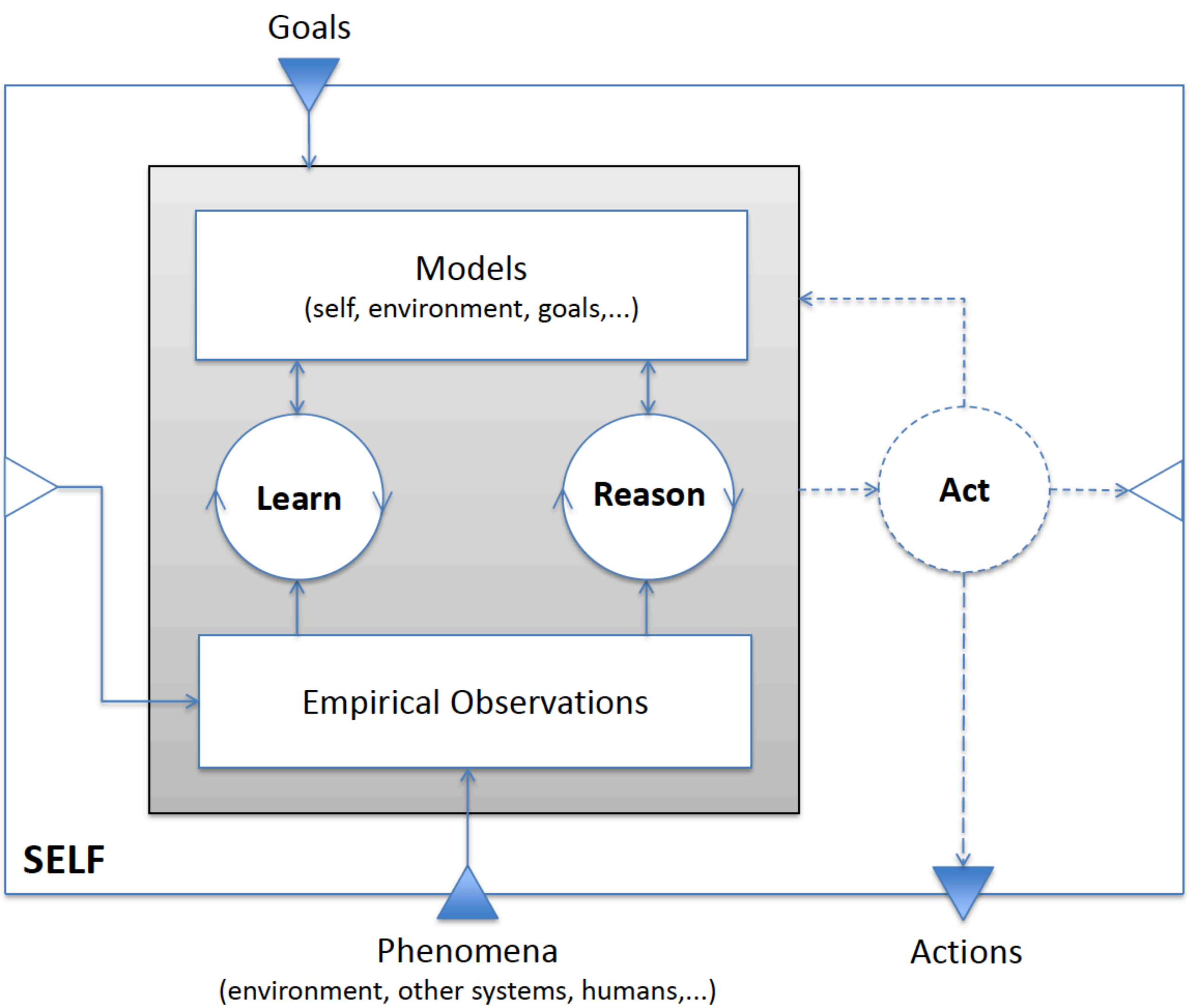
Self-aware Computing Systems are computing systems that:

1. ***learn models*** capturing ***knowledge*** about themselves and their environment (such as their structure, design, state, possible actions, and run-time behavior)
   on an ongoing basis and

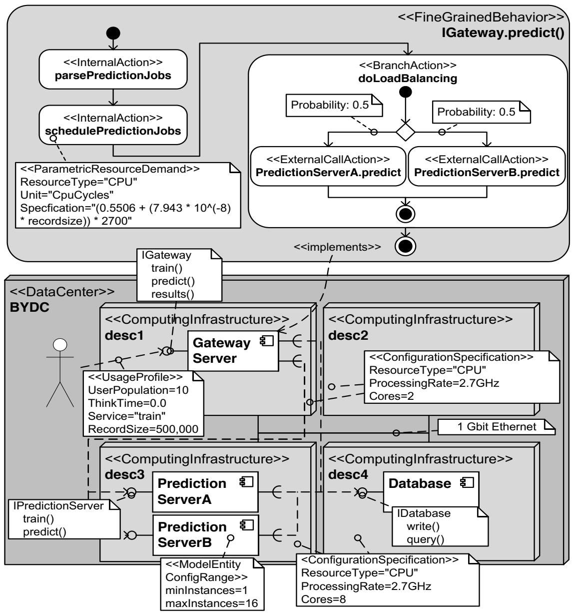
2. ***reason*** using the models (for example predict, analyze, consider, plan) enabling them to ***act*** based on their knowledge and reasoning (for example explore, explain, report, suggest, self-adapt, or impact their environment)

in accordance with ***higher-level goals***, which may also be subject to change.

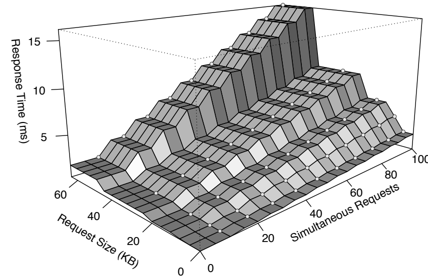
Self-Aware Learning & Reasoning Loop



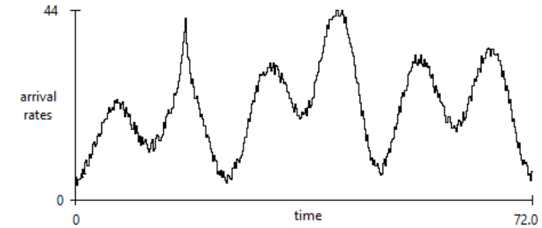
Examples of Models



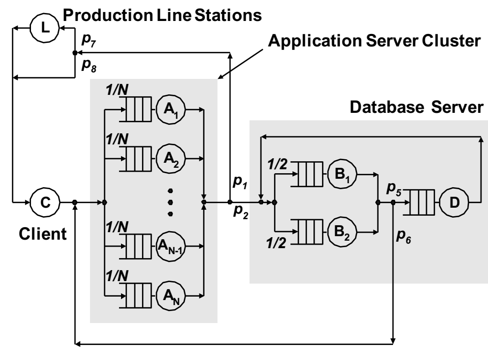
Descriptive MOF-based models



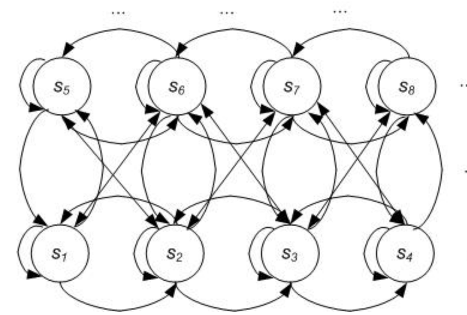
Statistical regression models



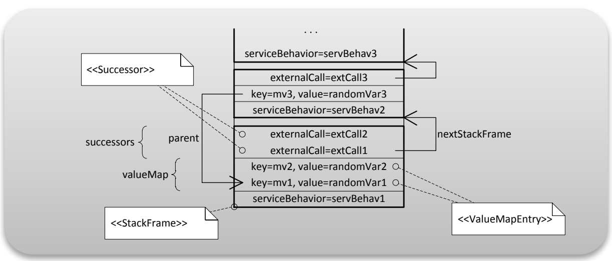
Load forecasting models



Queueing network models



Markov models



Simulation models

$$R \geq \max\left[N \times \max\{D_i\}, \sum_{i=1}^{K} D_i \right] \qquad X_0 \leq \min\left[\frac{1}{\max\{D_i\}}, \frac{N}{\sum_{i=1}^{K} D_i} \right]$$

$$\frac{N}{\max\{D_i\}[K+N-1]} \leq X_0 \leq \frac{N}{avg\{D_i\}[K+N-1]}$$

Analytical analysis models

Telescope: A Hybrid Forecast Method for Univariate Time Series

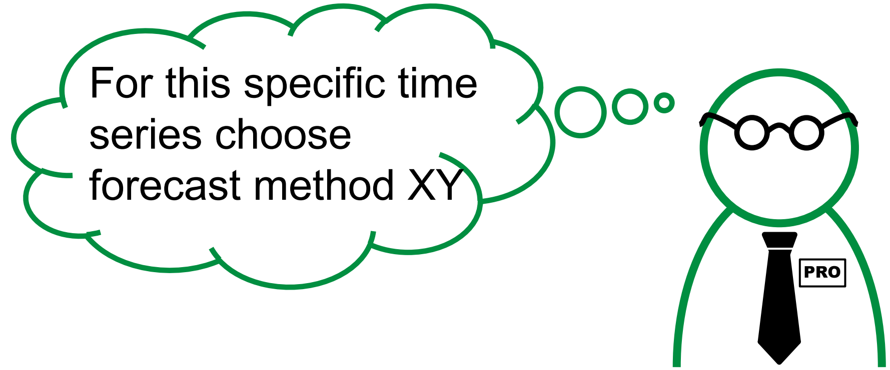
Marwin Züfle, N. Herbst, A. Bauer, V. Curtef, S. Kounev

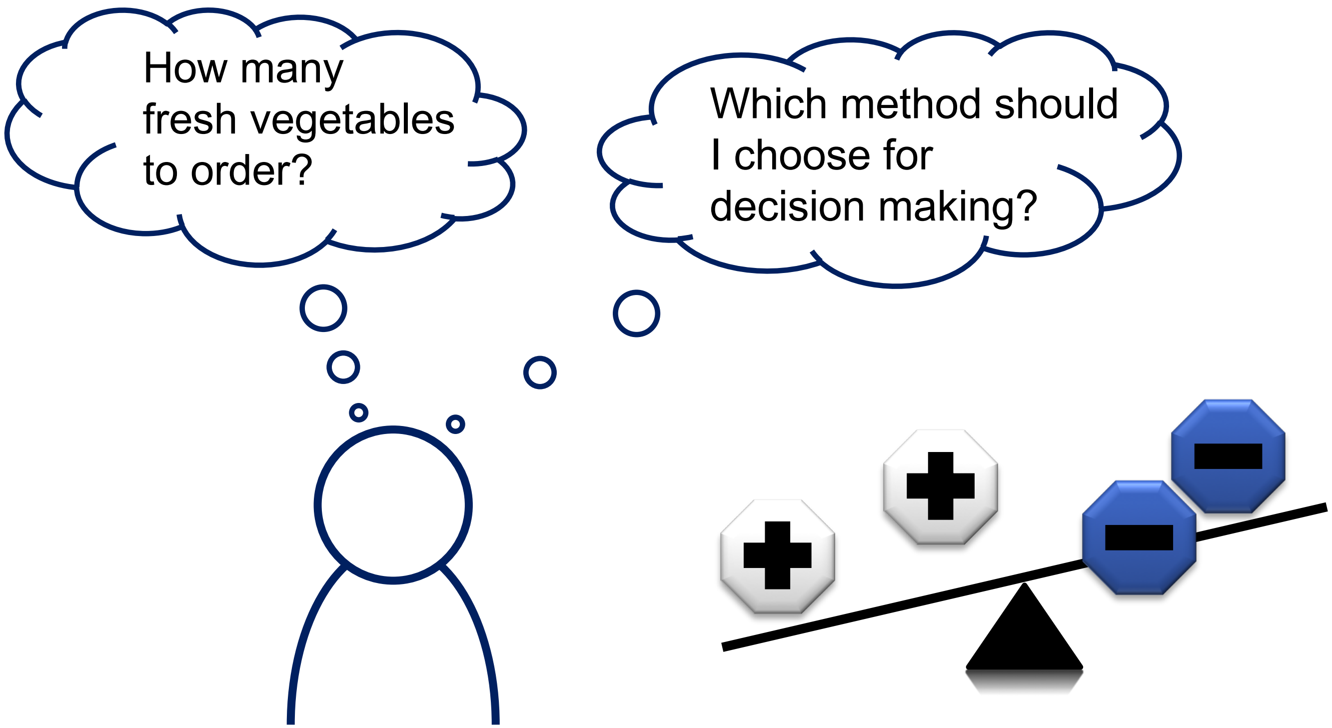
Chair of Software Engineering
Descartes Research Group
University of Würzburg

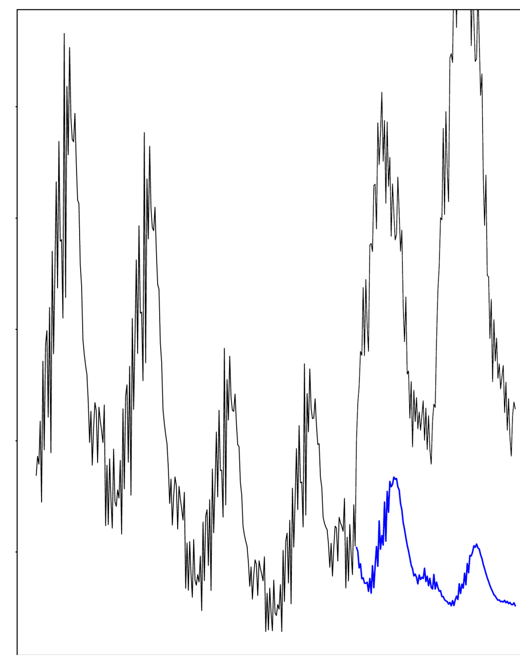
MaxCon Data Science GmbH

Slides available: http://descartes.tools/telescope

Forecasting Motivation

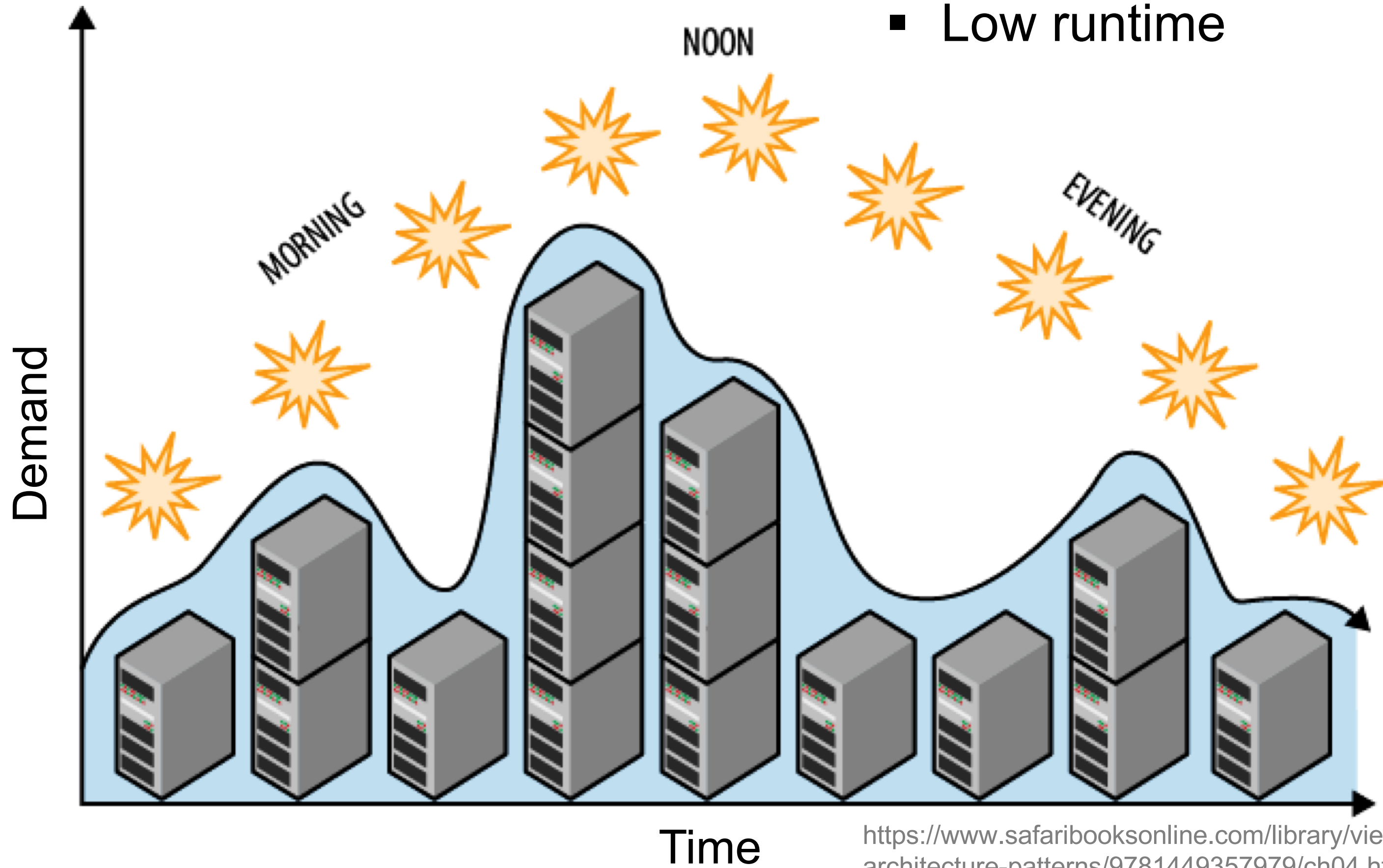






Use Case: Proactive Auto-Scaling

- Properties:
  - Seasonal time series
  - Short measurement intervals
- Requirements:
  - High accuracy
  - Stable multi-step forecaster
  - Low runtime



https://www.safaribooksonline.com/library/view/cloud-architecture-patterns/9781449357979/ch04.html

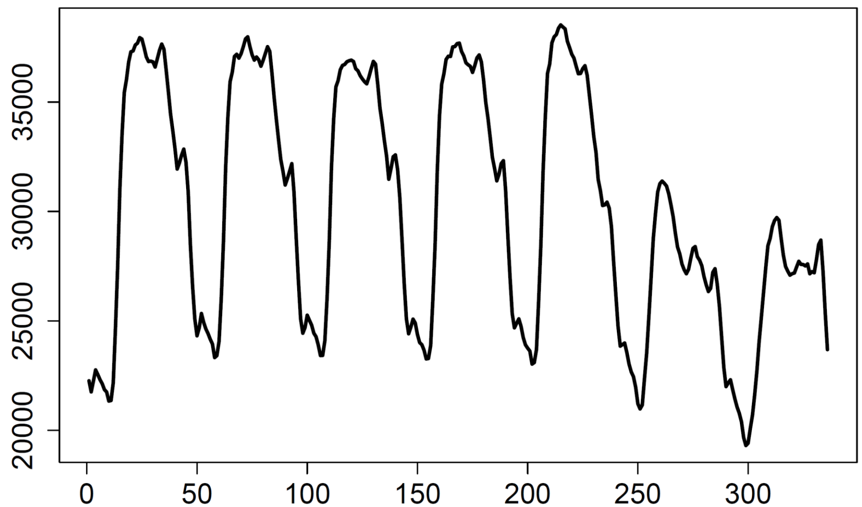
IoT / Industry 4.0

Use Case: IoT Sensor Data

Temperature

Humidity

Acceleration



Distance

Oscillation

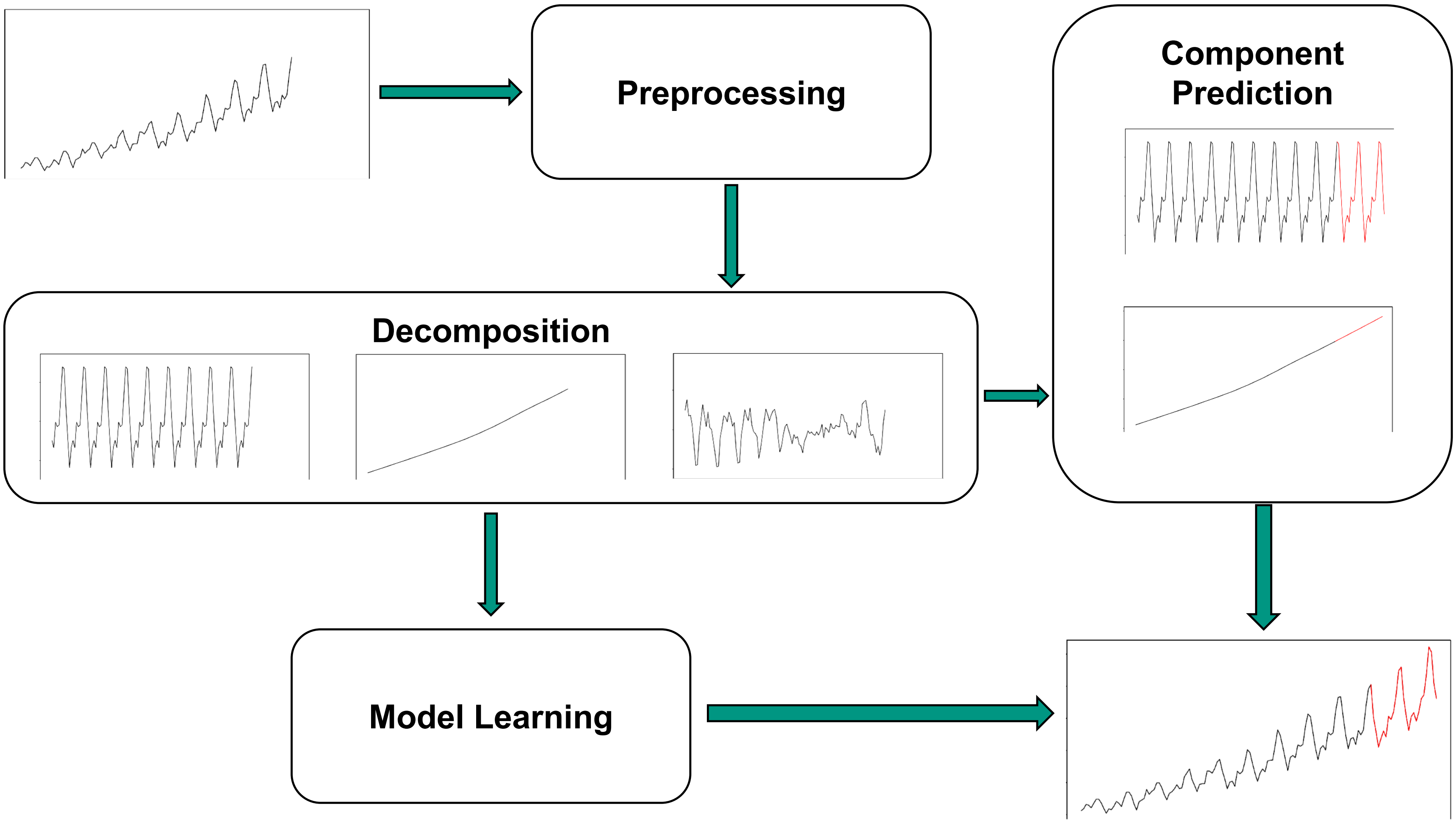
Light

Pressure

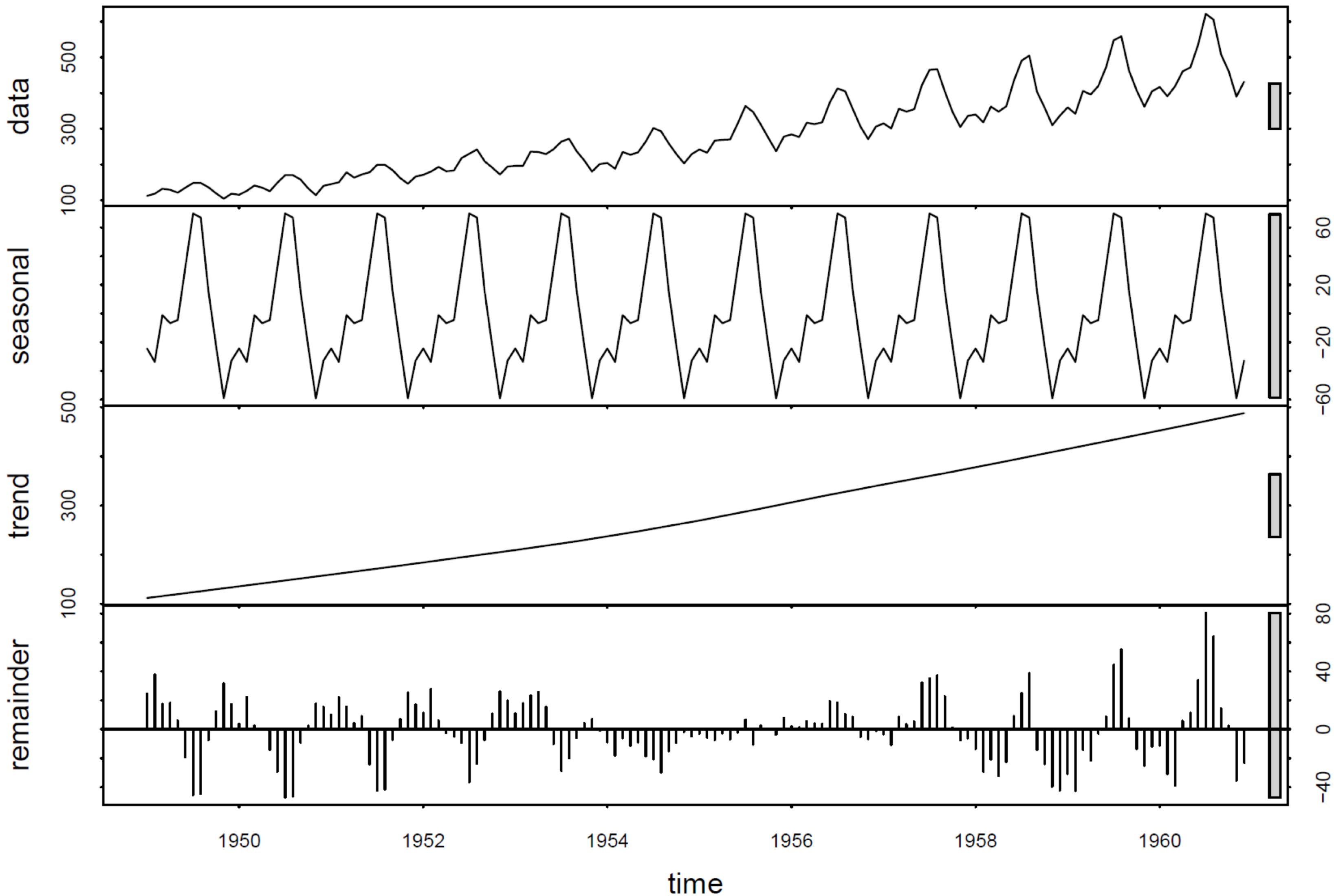
Speed

Inclination

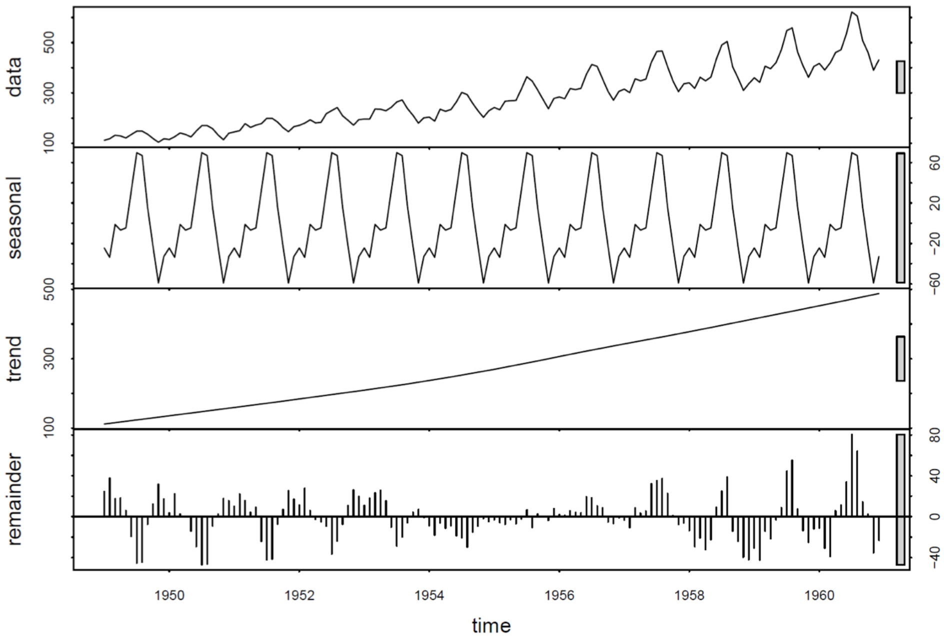
Forecasting Approach



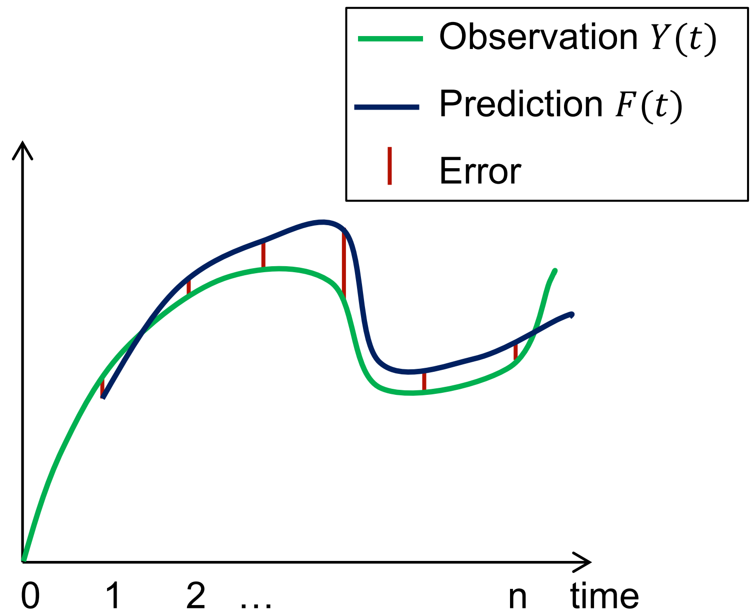
Forecasting Time Series



Forecasting Time Series



$$Y(t) = S(t) \diamond T(t) \diamond I(t)$$



$$E(t) = |F(t) - Y(t)|$$

Related Work – Hybrid Forecasting

Ensemble Forecasting

- Historically first hybrid forecasting task
- Weighting results from several methods
- Linear combination of these weighted results

[*Bates69, Clemen89, Menezes00*]

Forecaster Recommendation

- Learning a rule set from a set of time series
- System guesses best forecasting method
- Expert system or machine learning techniques

[*Collopy92, Wang09*]

Decomposition Forecasting

- Combining advantages of different methods
- Decomposition of time series into its components
- Executing several methods one after another

[*Zhang03, Pai05, Liu14, Züfle17*]

In a Nutshell

Motivation

- Use case Auto-Scaling: forecasting results required within a fixed time slot
- Multi-step-ahead forecasting
- *No-Free-Lunch* theorem

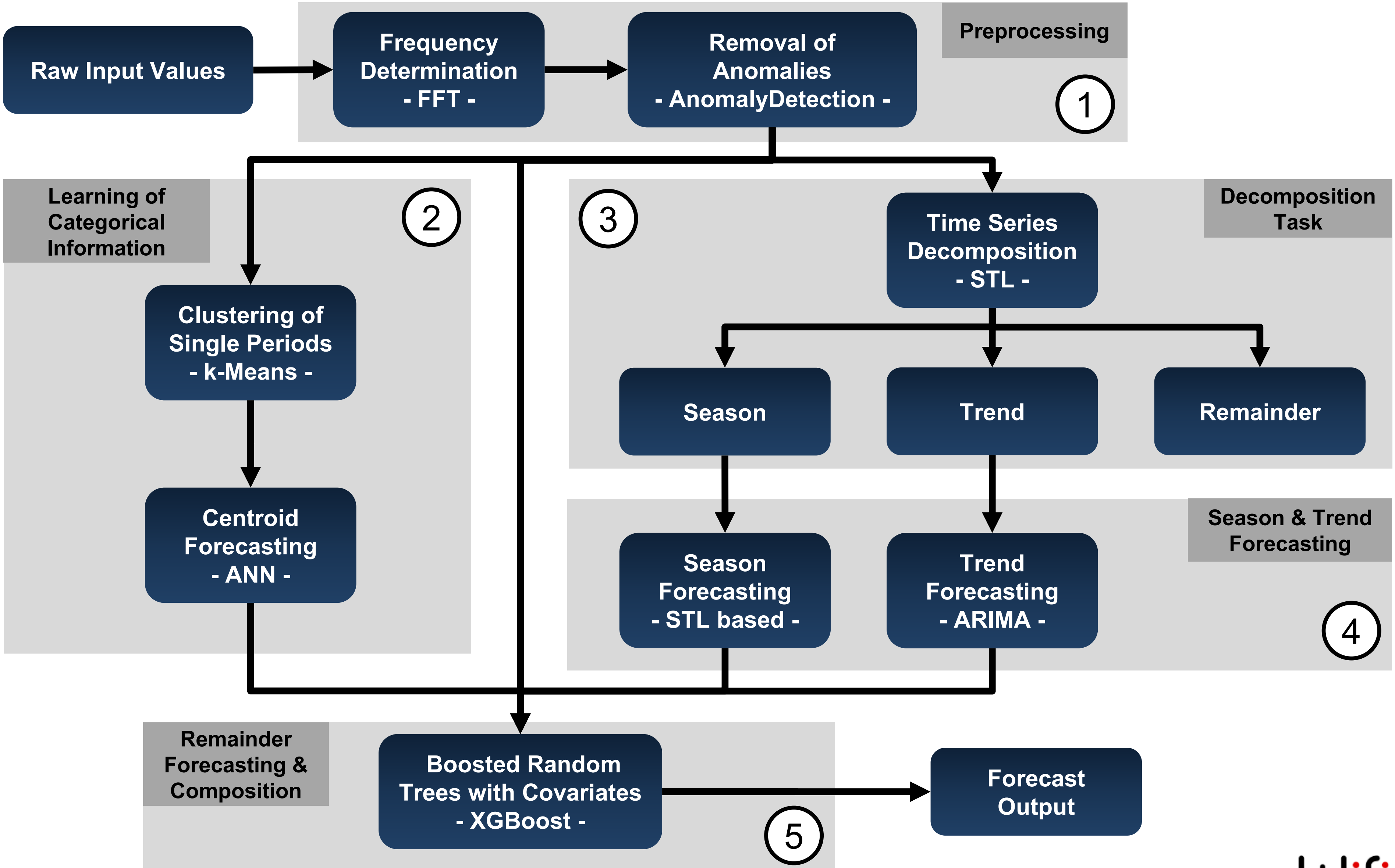
Idea

- Preprocessing (frequency, anomaly)
- Decomposition and composition
- Learning of categorical information
- Specific forecasting method per component

Benefits

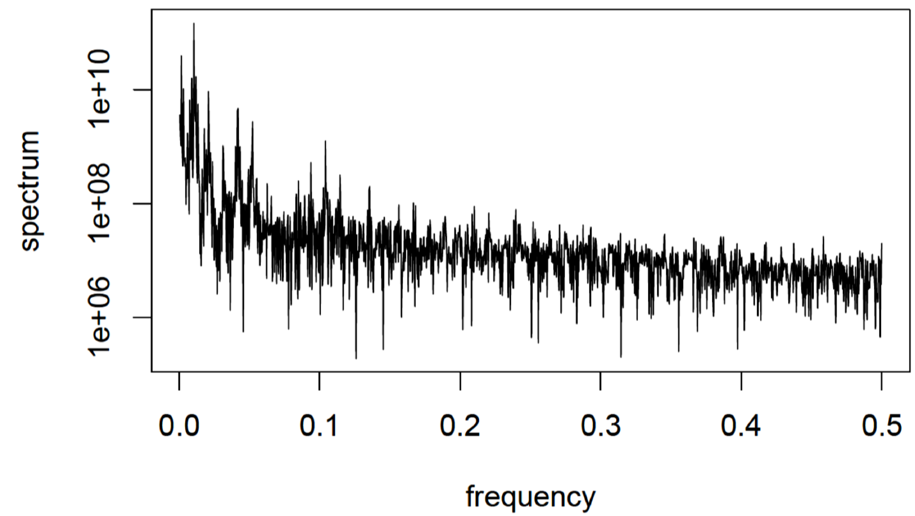
- Improvement of forecast accuracy
- Variance reduction for forecasting results
- Short runtime

Telescope Approach

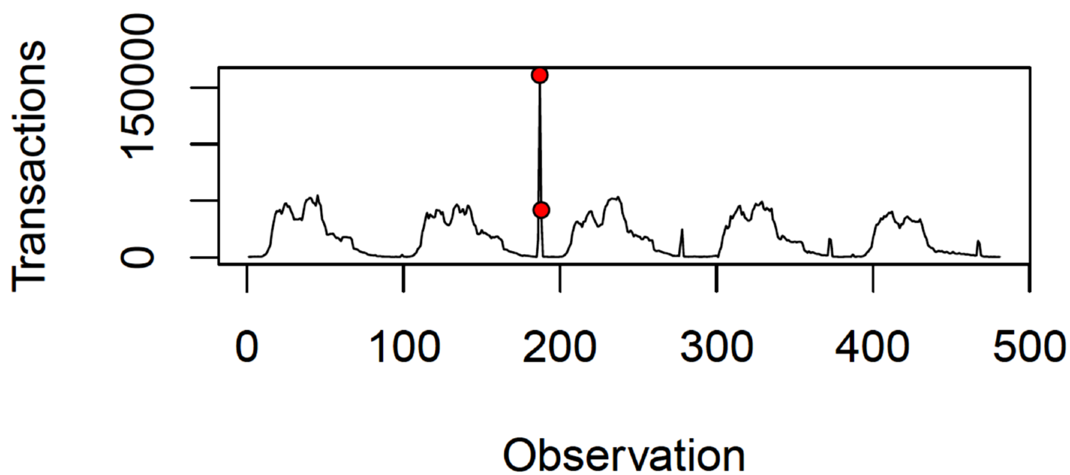


① Preprocessing

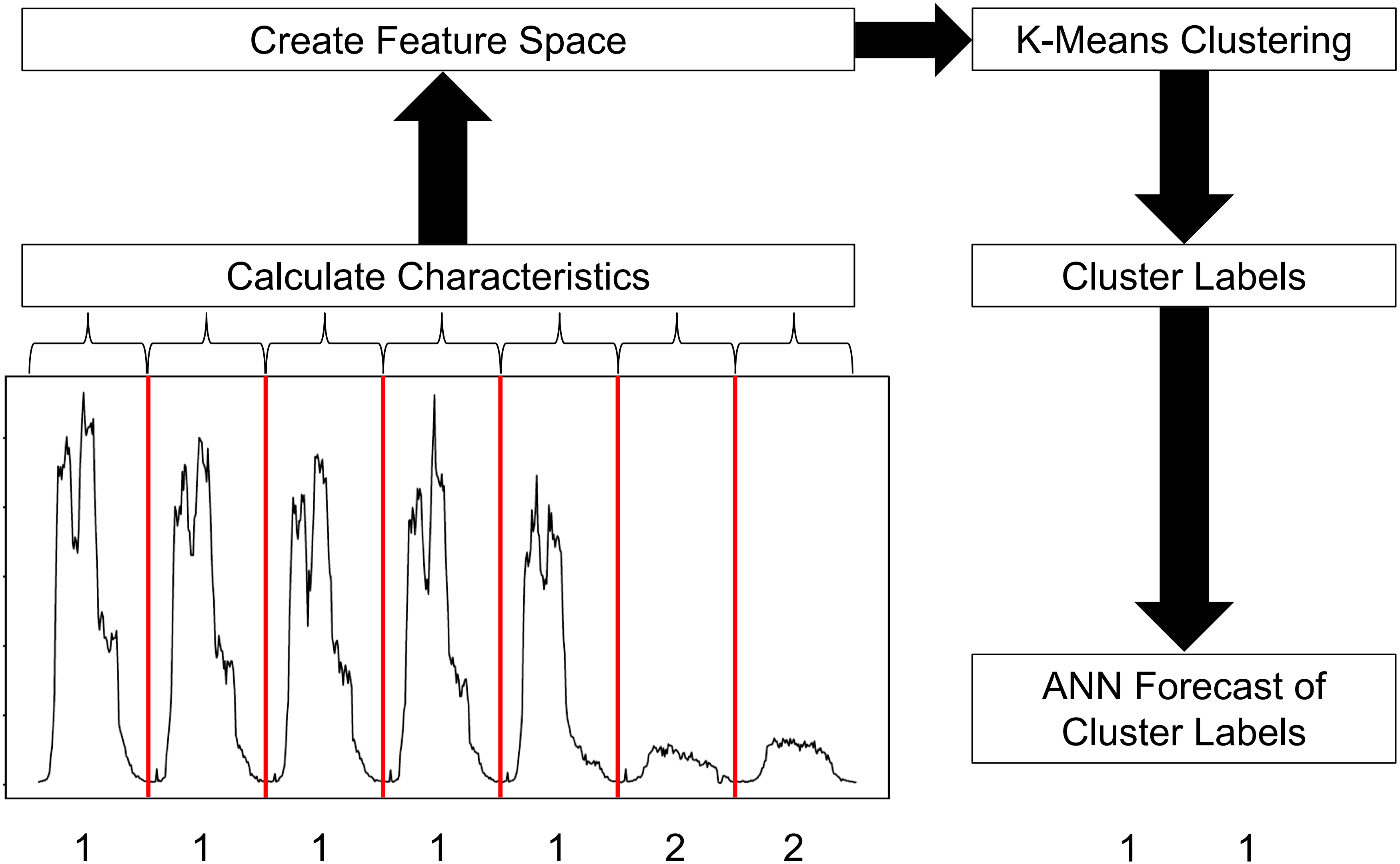
- Frequency Estimation:
  - Periodograms for rough estimation
  - List of common frequencies



- Anomaly Detection:
  - Generalized extreme studentized deviate test (ESD) on the remainder
  - Replace anomaly by mean of non-anomaly neighbors



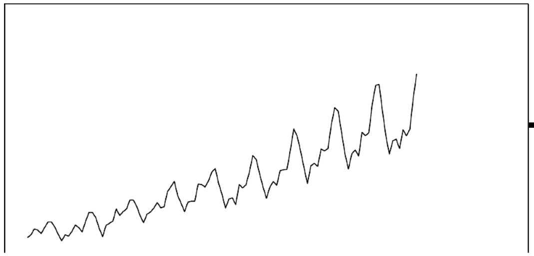
② Learning Categorical Information



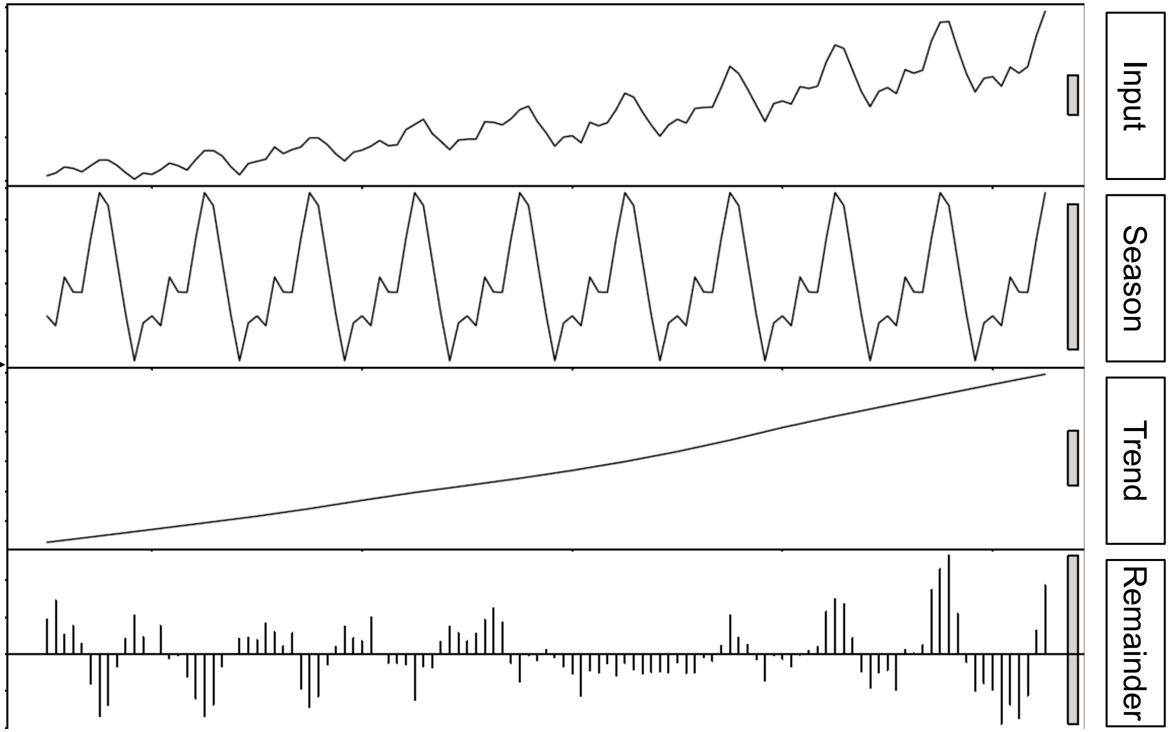
3 4 5 Decomposition & Forecasting

Learning of Categorical Information

Cluster Label Forecast



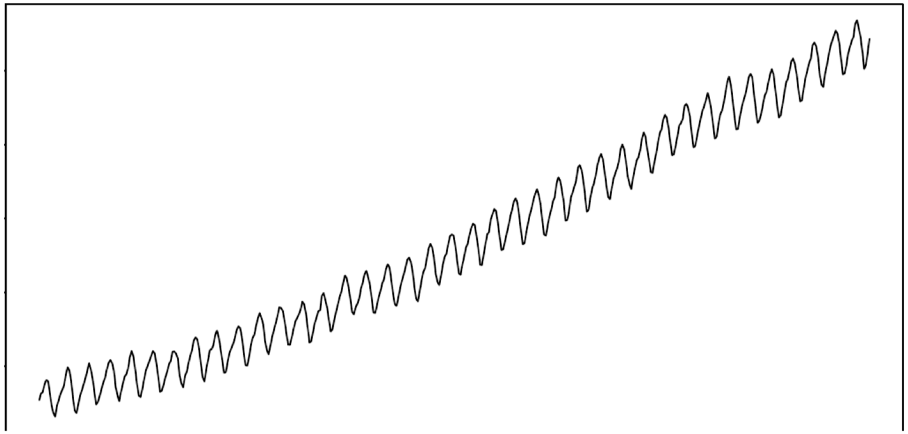
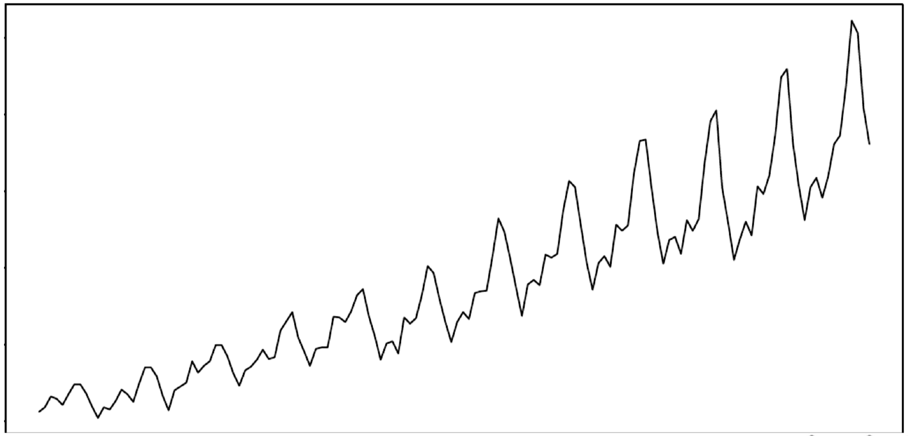
Time Series History



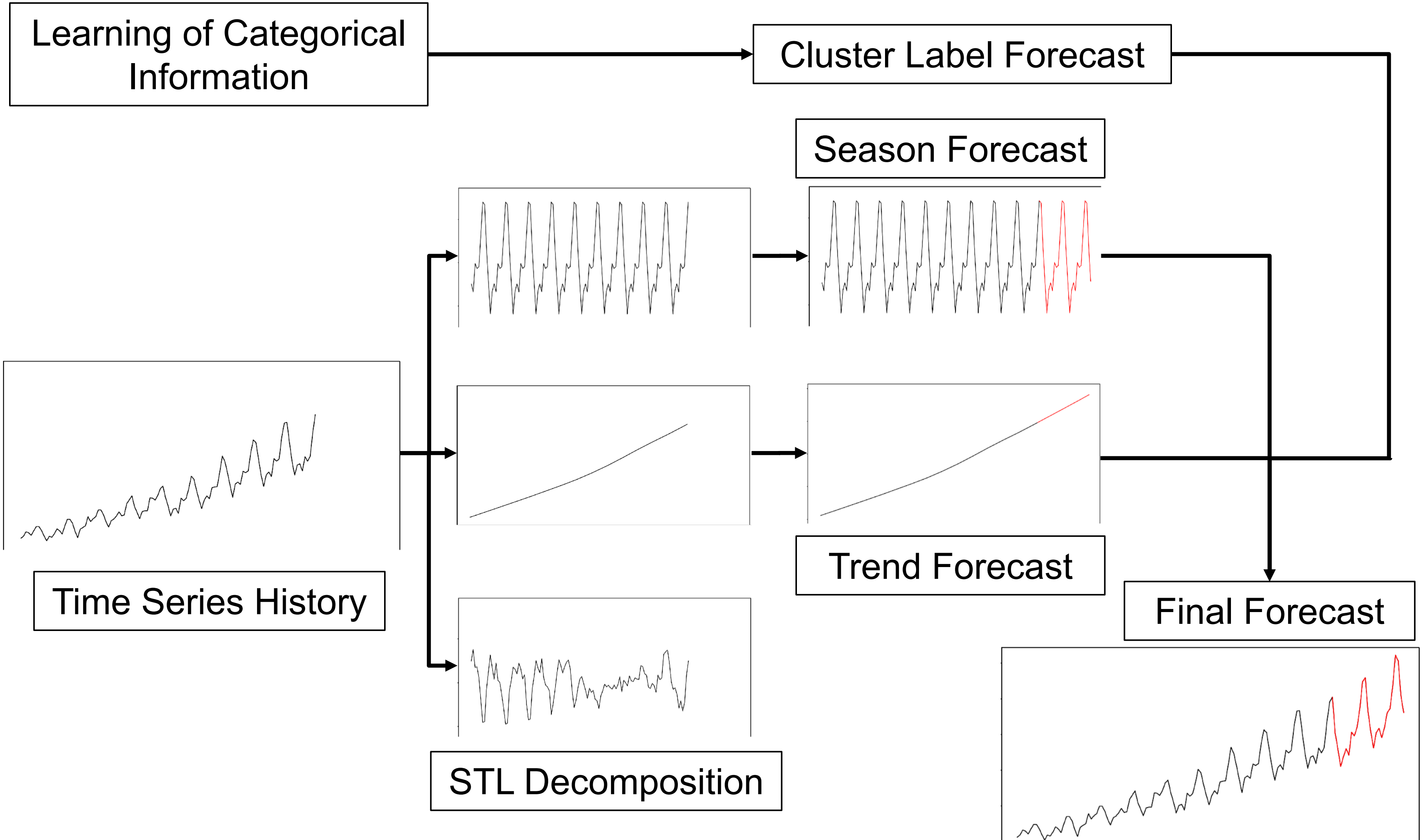
STL Decomposition

③ Estimating Decomposition Type

- STL once on original and once on logarithmized time series
- Calculate:
  - Sum of squares of the auto-correlation on remainder
  - Range between first and third quantile of the remainder
  - Sum of squares of the remainder
- Majority decision

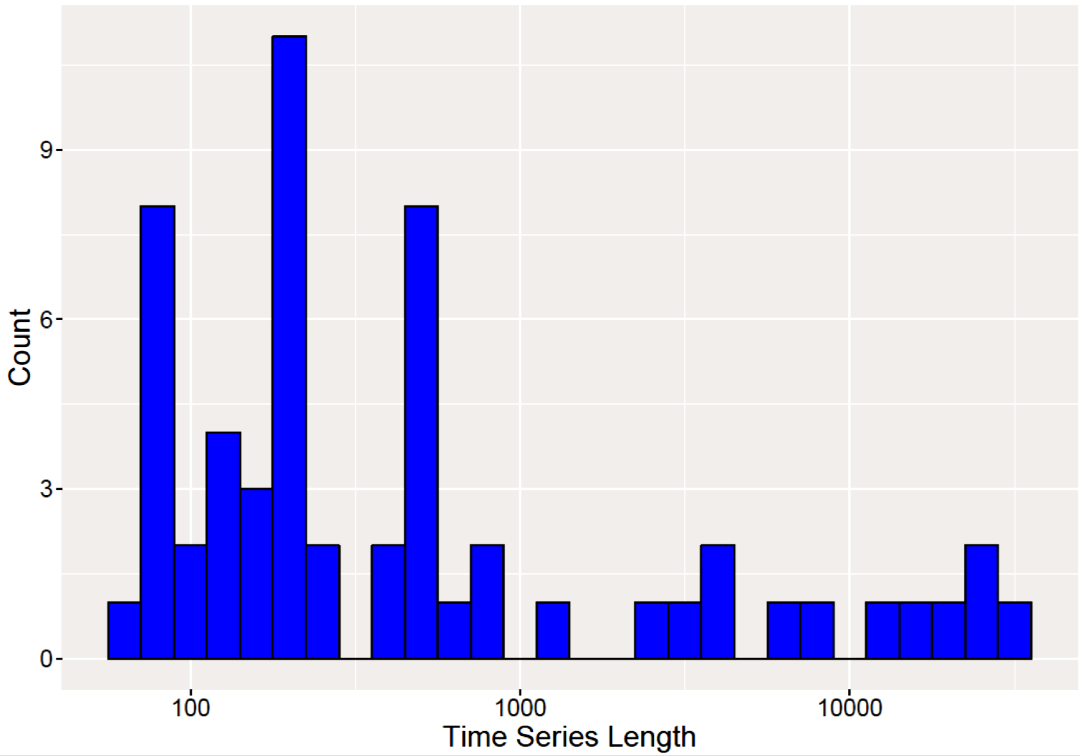



Decomposition & Forecasting



Evaluation Design – Data Set

- Data Sets:
  - 56 seasonal time series
  - Mostly short time series
  - Last 20% as forecasting horizon



Sources:

- Internet Traffic Archive
- Wikipedia project-counts
- German federal statistical office
- R packages

  forecast, **fpp2**, **datasets**

Telescope

http://descartes.tools/telescope

Evaluation Design

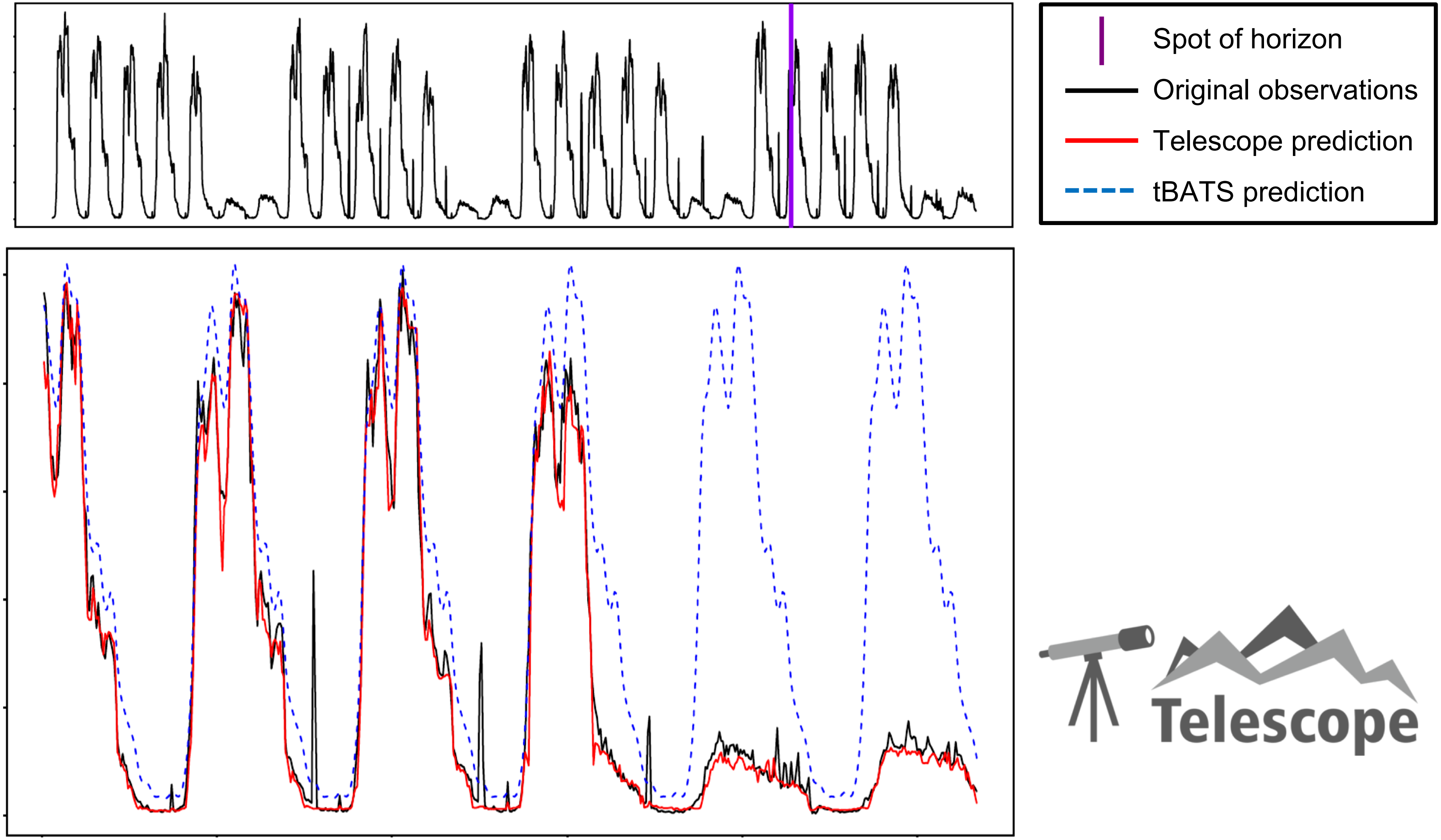
- Methods in Competition:
  - (S)ARIMA
  - ETS
  - tBATS
  - ANN
  - SVM
  - XGBoost

http://descartes.tools/telescope

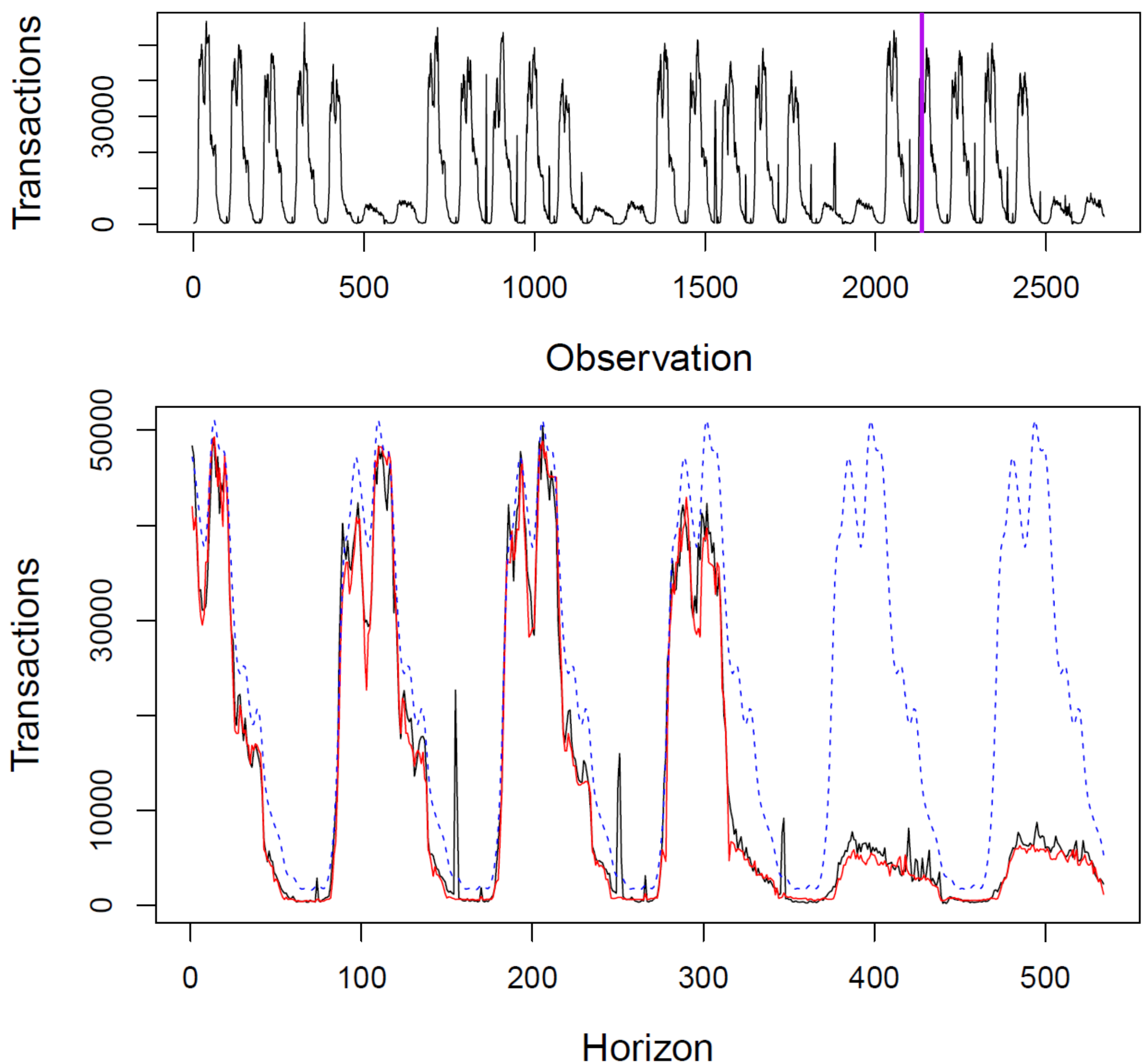
- Evaluation Measures:
  - $MASE = \dfrac{\frac{1}{n} \times \sum_{t=1}^{n} |e_t|}{\frac{1}{n-1} \times \sum_{t=2}^{n} |Y_t - Y_{t-1}|}$
  - $MAPE = 100\% \times \frac{1}{n} \sum_{t=1}^{n} \left| \frac{e_t}{Y_t} \right|$
  - Time-to-result [secs]

n: Length of the forecasting horizon
Y_t: Real value at time t
F_t: Forecast at time t
$e_t = Y_t - F_t$: Forecasting error at time t

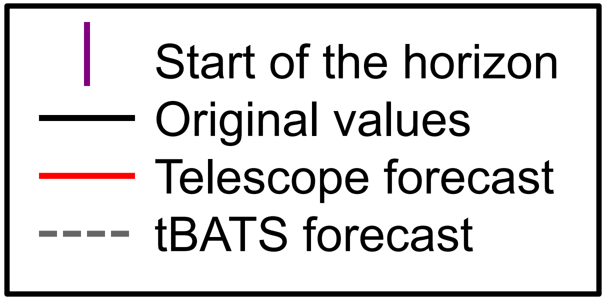
Forecasting Approach Evaluation



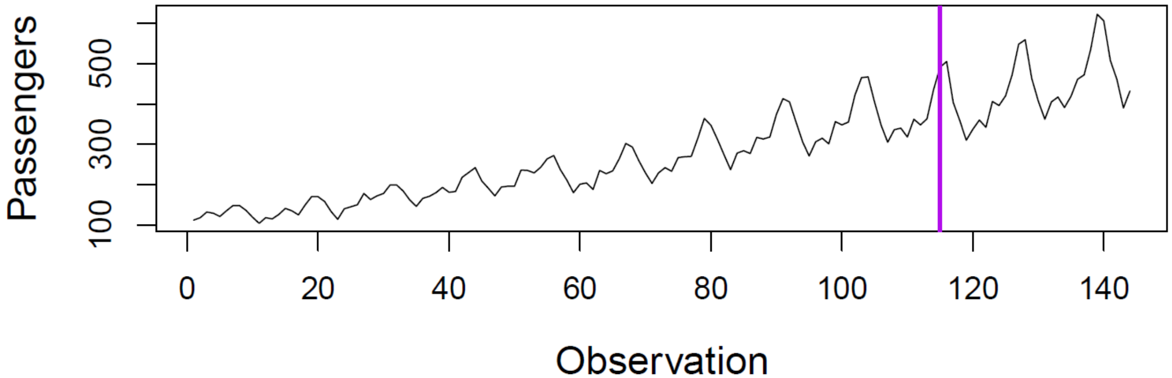
Example: IBM Trace

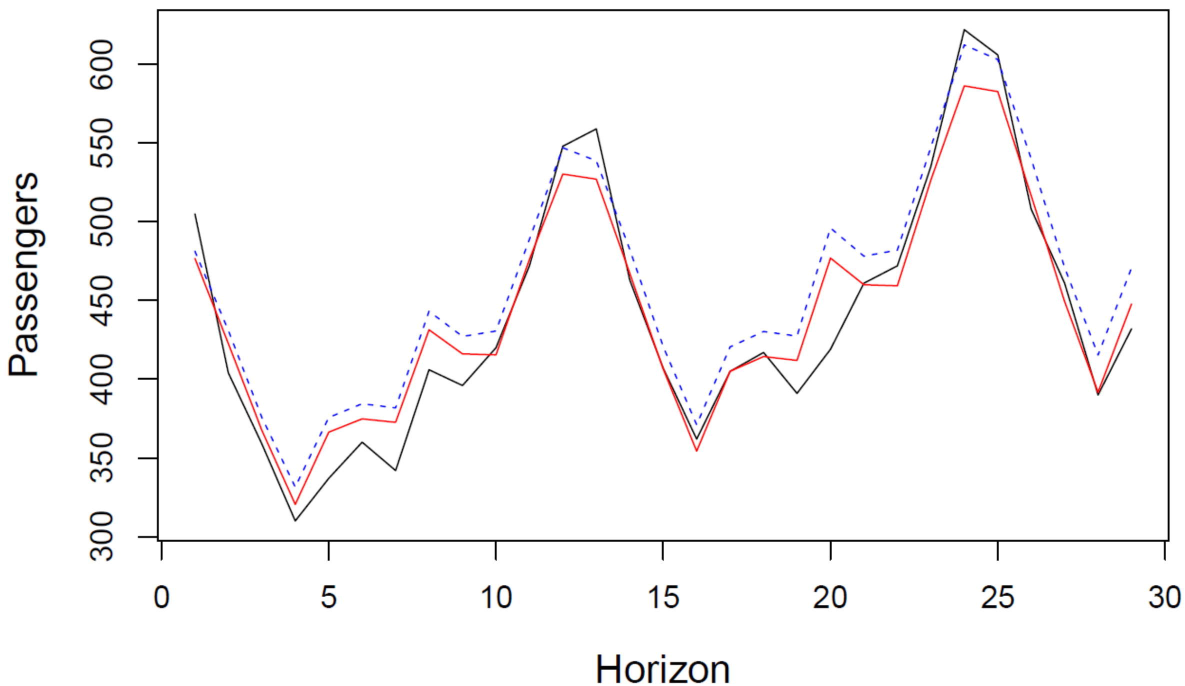


Forecaster	MASE	Time
Telescope	**0.842**	6.248
tBATS	4.547	33.360
SVM	6.557	2.344
XGBoost	7.683	**0.172**
ARIMA	7.828	87.016
ANN	18.678	10.938
ETS	23.389	0.984

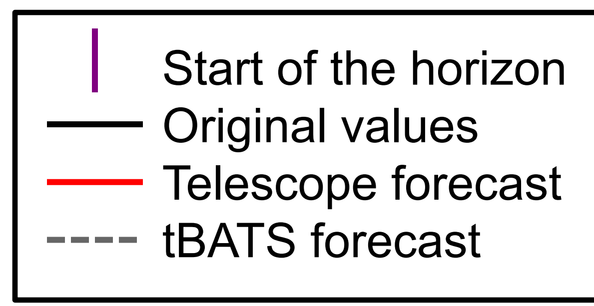


Example: Airline Passengers Trace





Forecaster	MASE	Time
Telescope	**0.353**	1.671
tBATS	0.520	11.641
ARIMA	0.638	3.248
ETS	0.652	2.266
ANN	0.711	0.375
XGBoost	1.261	0.102
SVM	6.758	**0.094**



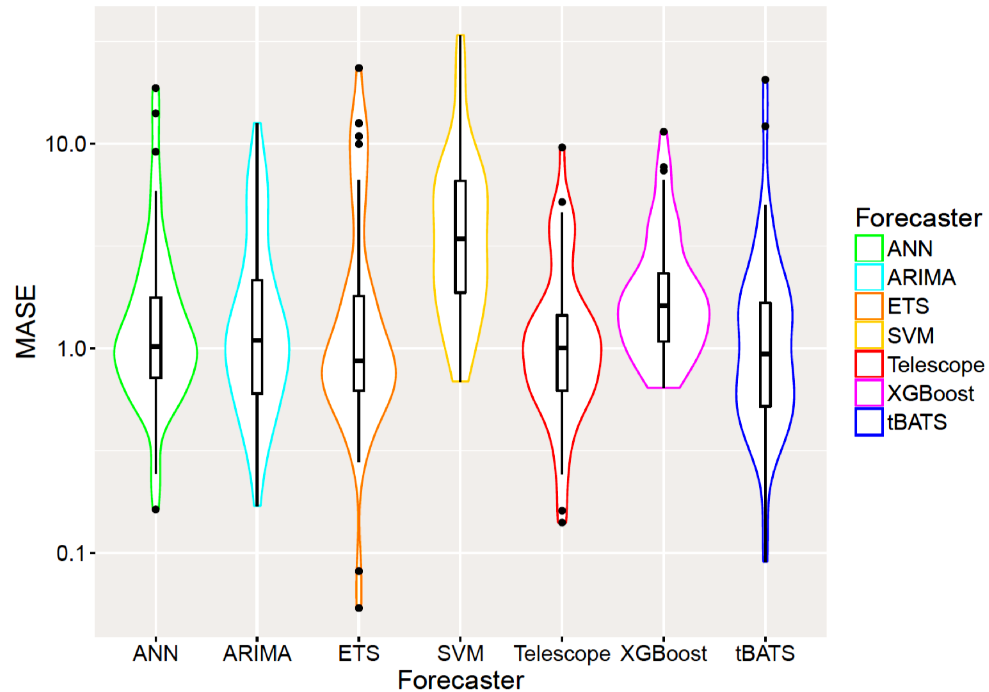
Measures for the 56 Time Series

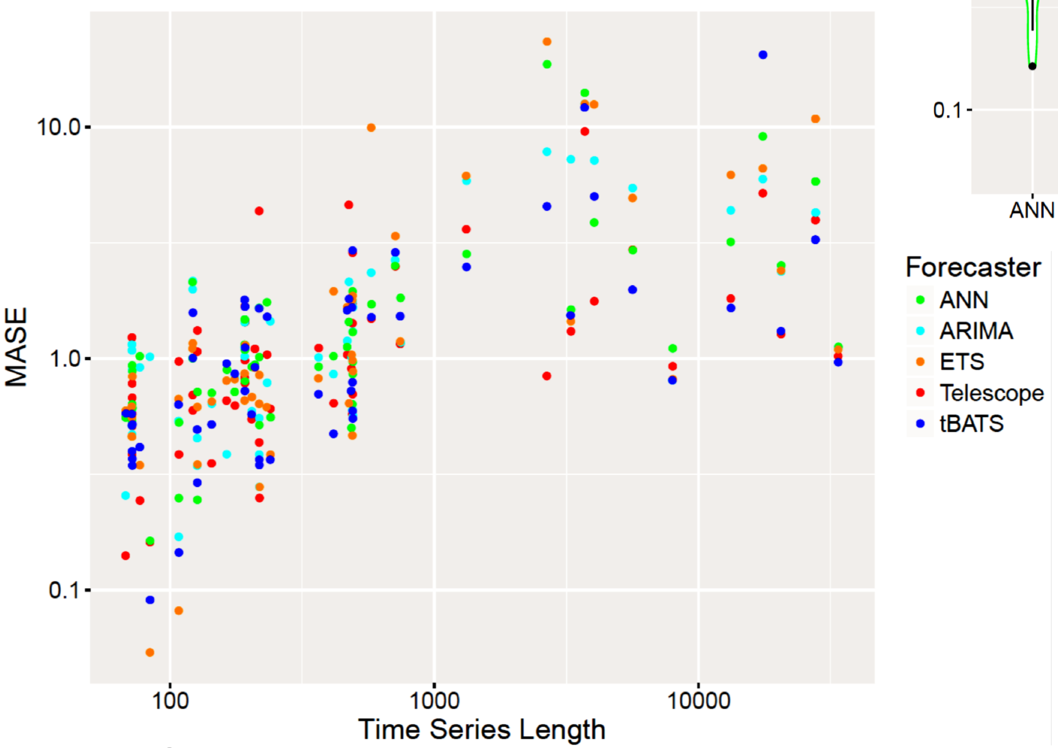
Forecaster	Ø MASE	σ MASE	Ø MAPE	Ø Time
Telescope	**1.503**	**1.619**	25.217	9.032
tBATS	1.791	3.112	**25.107**	56.334
ARIMA	2.022	2.405	43.194	177.288
ANN	2.072	3.206	67.176	77.948
XGBoost	2.251	2.017	47.779	**0.167**
ETS	2.638	4.288	81.816	2.184
SVM	5.334	6.254	64.306	24.608

- High and stable accuracy for multi-step forecasting
- Comparably short time-to-results

Variance of Accuracy for the Entire Data Set

- Variance reduction for accuracy
- Accuracy less effected by time series length

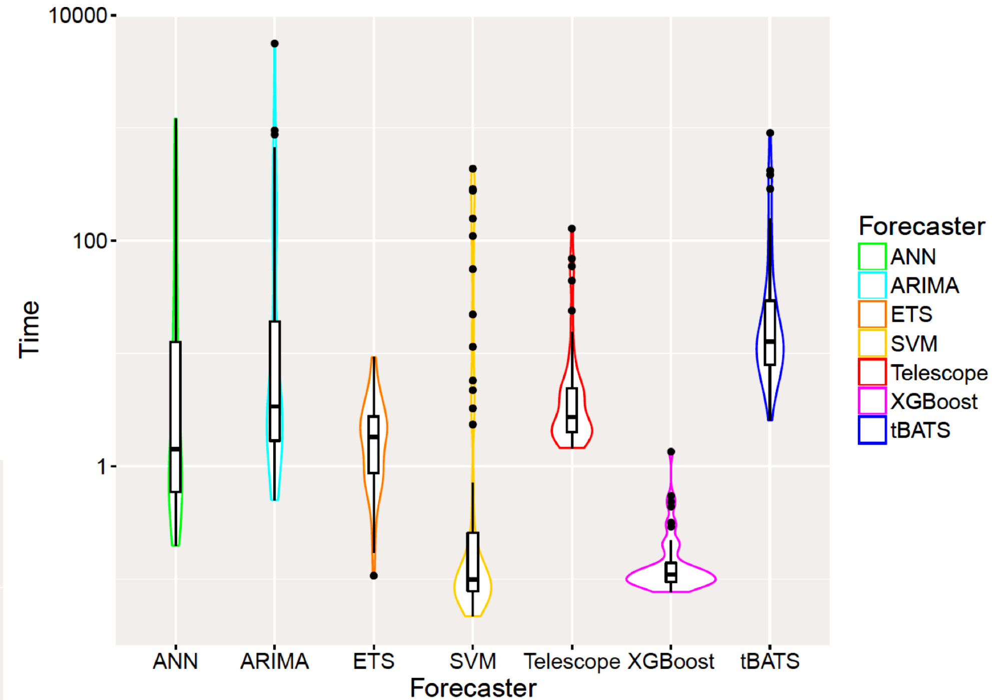


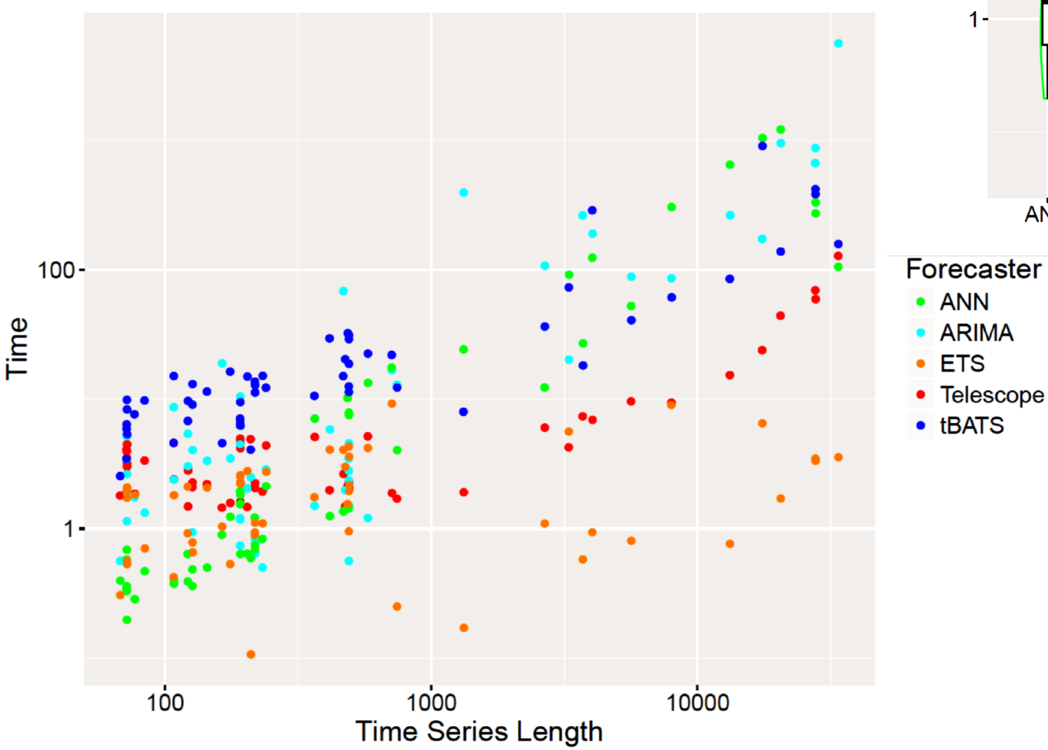


Forecaster	σ MASE
Telescope	**1.619**
ANN	3.206
ARIMA	2.405
ETS	4.288
tBATS	3.112

Variance of Time-to-Result for the Entire Data Set

- Low variance in runtime
- Runtime less effected by time series length





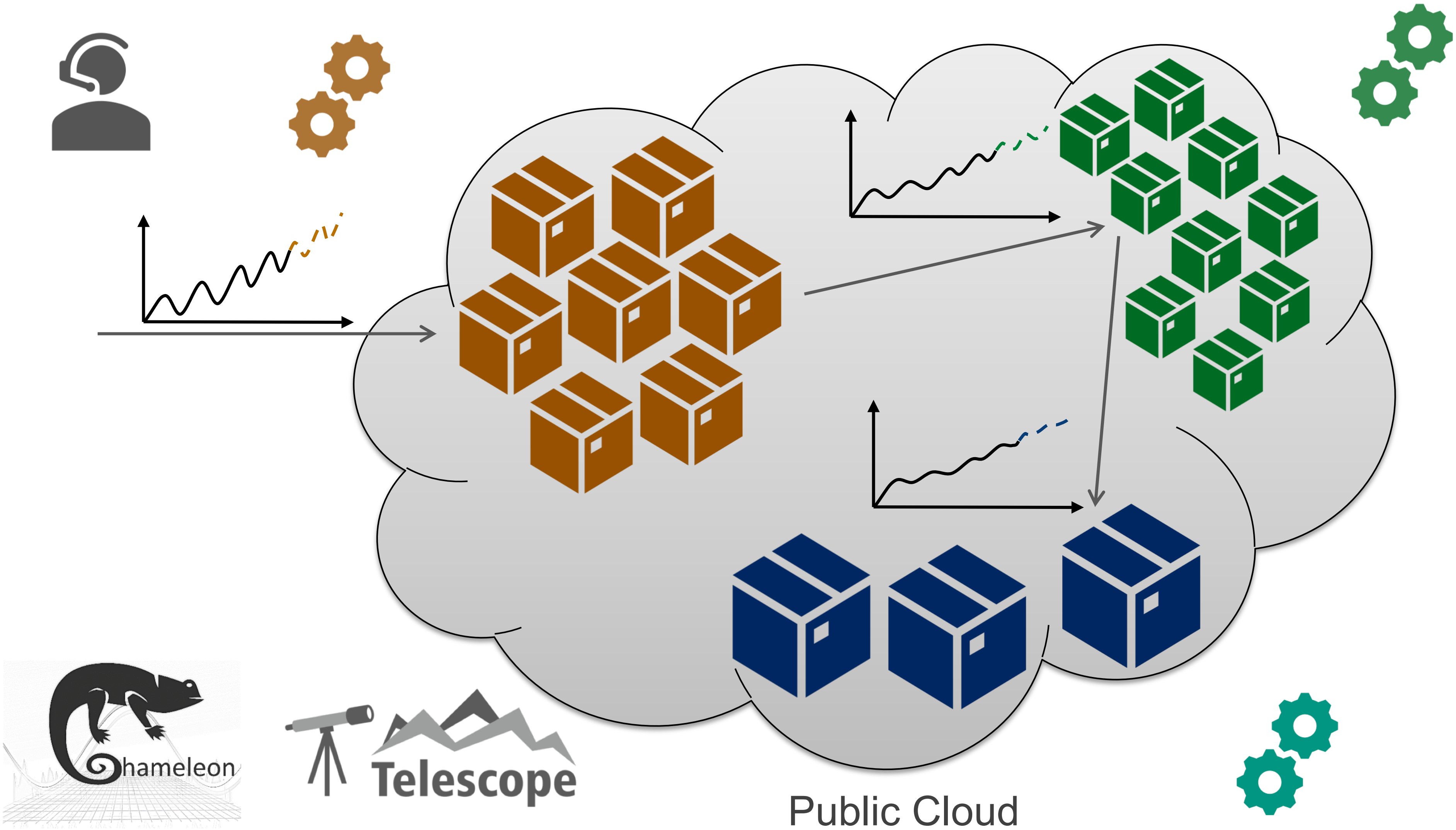
Forecaster	σ Time
Telescope	20.760
ANN	232.150
ARIMA	766.829
ETS	**1.934**
tBATS	142.740

Measures for Long Time Series

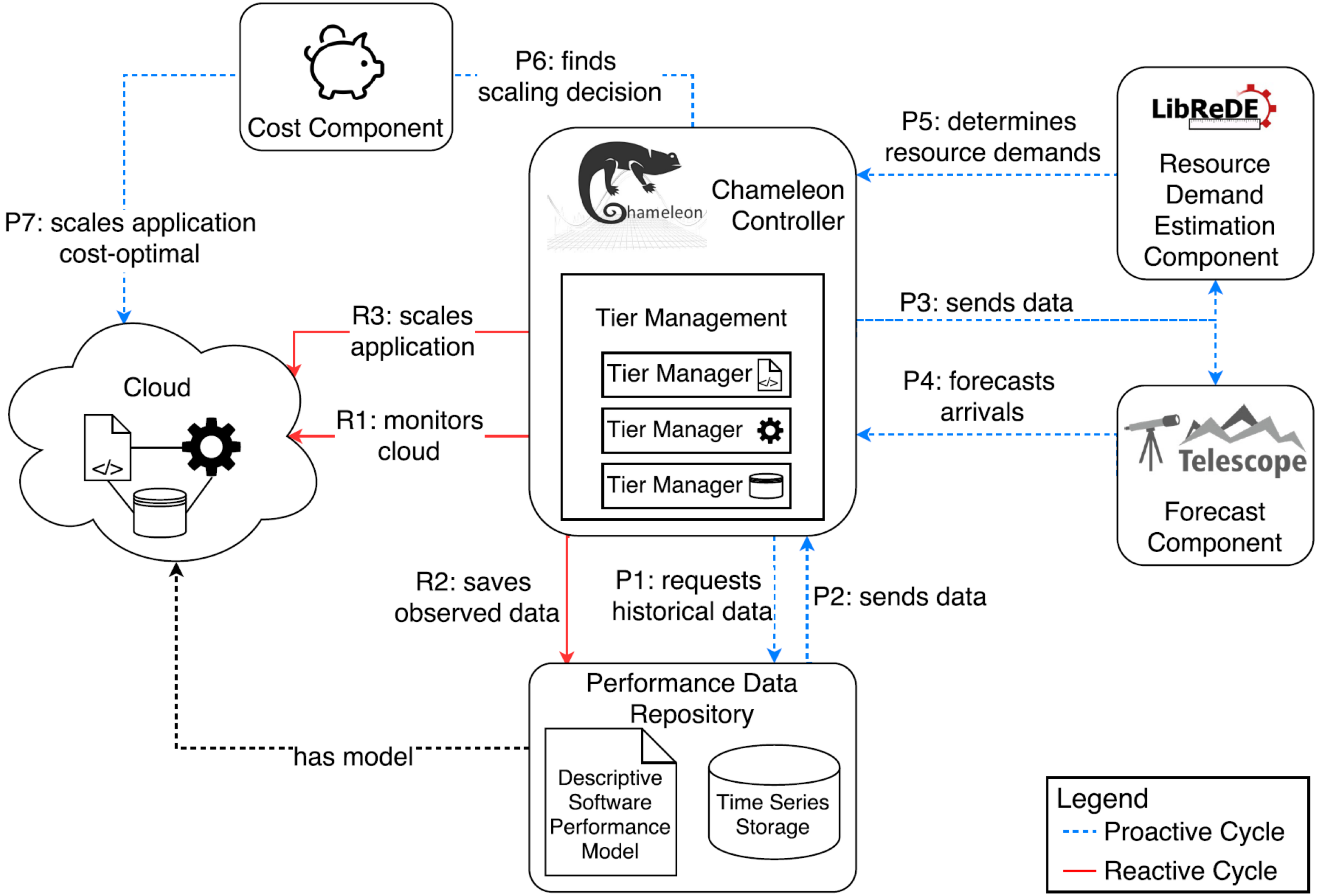
Forecaster	Ø MASE	Ø MAPE	Ø Time
Telescope	**2.823**	**47.218**	23.313
tBATS	3.955	54.503	158.464
ARIMA	4.569	102.783	572.796
ANN	4.719	190.557	252.470
XGBoost	3.938	100.147	**0.308**
ETS	6.769	241.223	3.196
SVM	5.817	93.263	80.854

- Highest and most stable accuracy for long time series
- Comparably short time-to-result

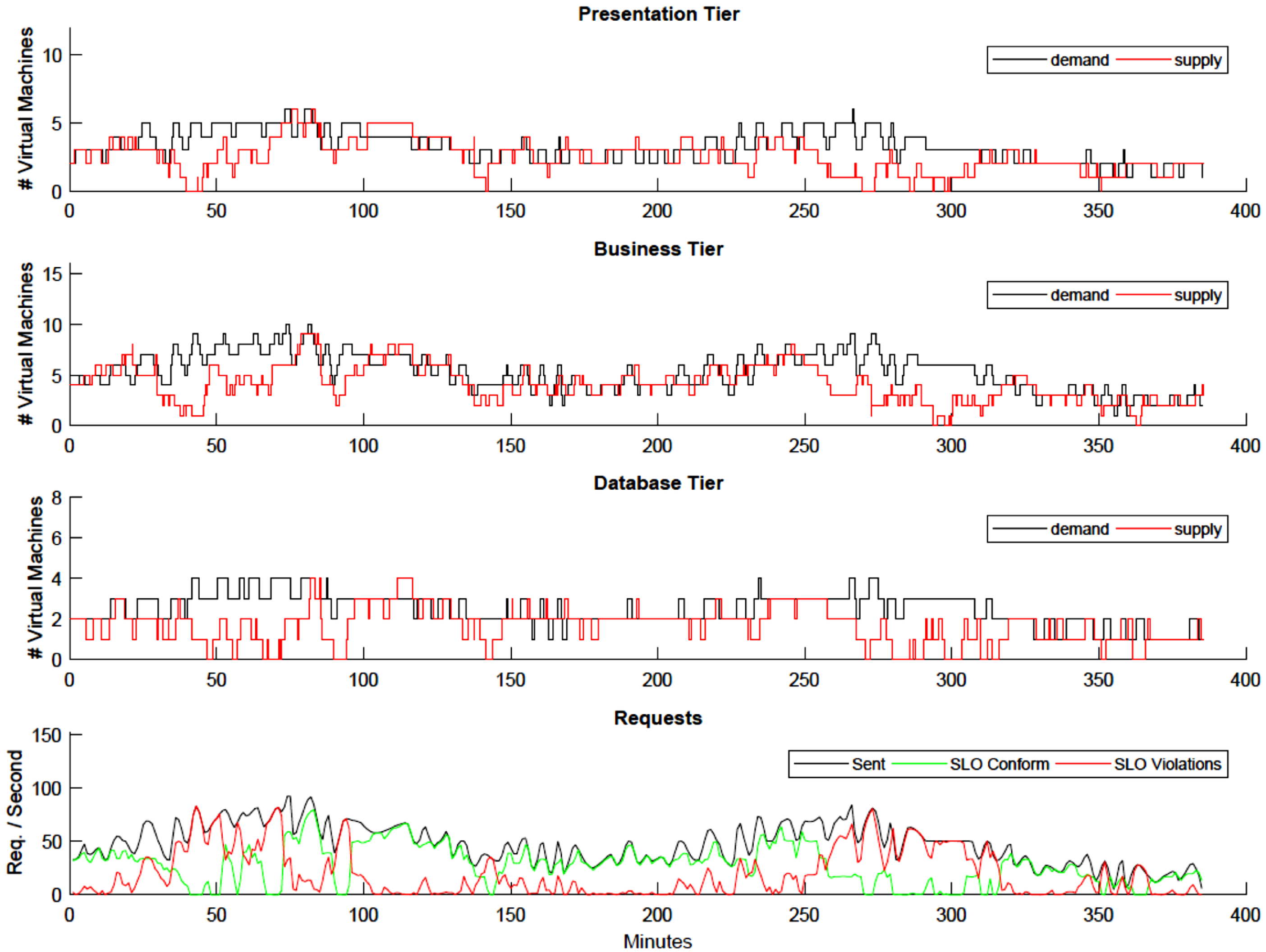
Multi-Tier Auto-Scaling



Chameleon Approach

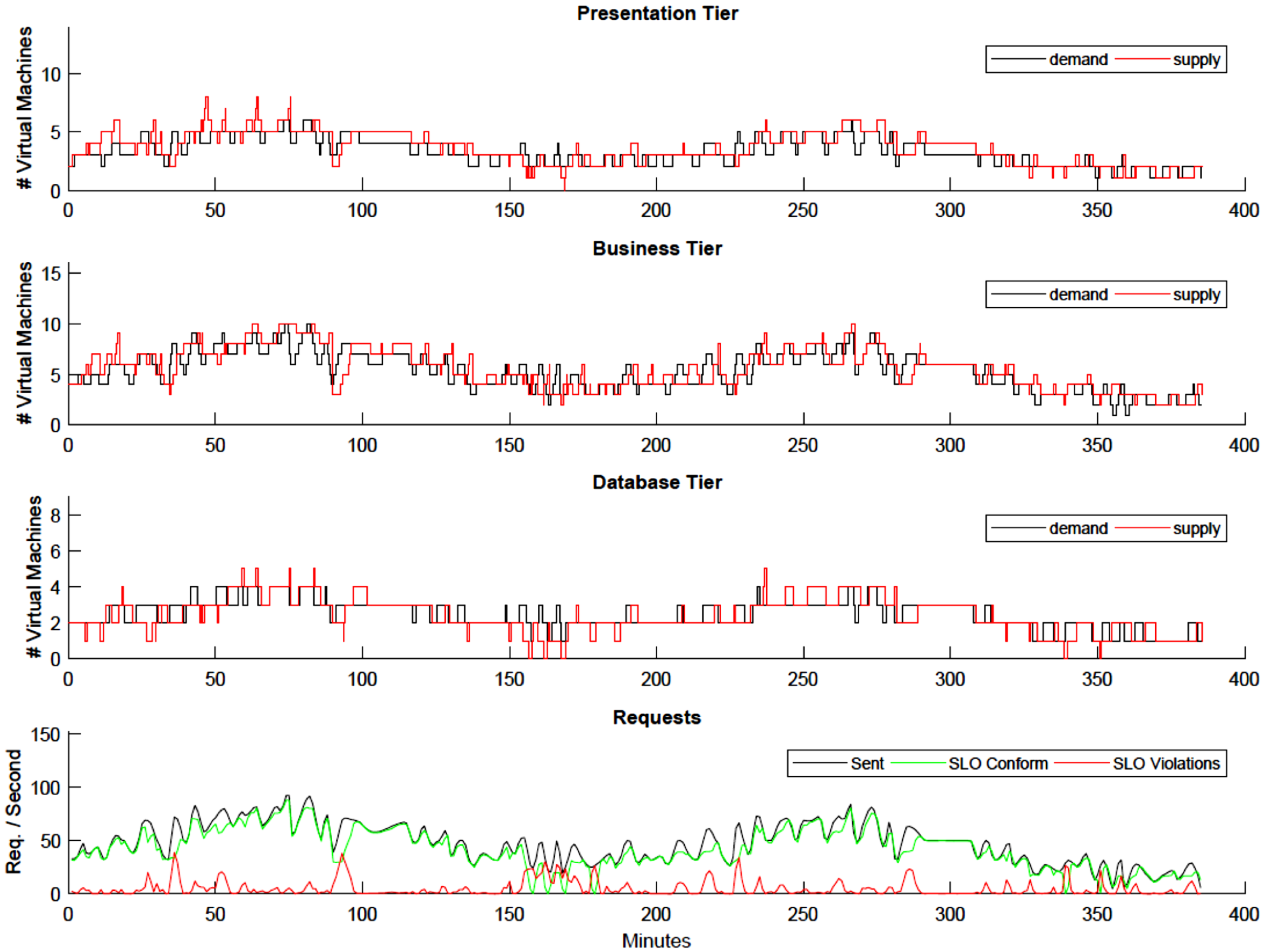


Auto-Scaling using tBATS



36% SLO violations

Auto-Scaling using Telescope

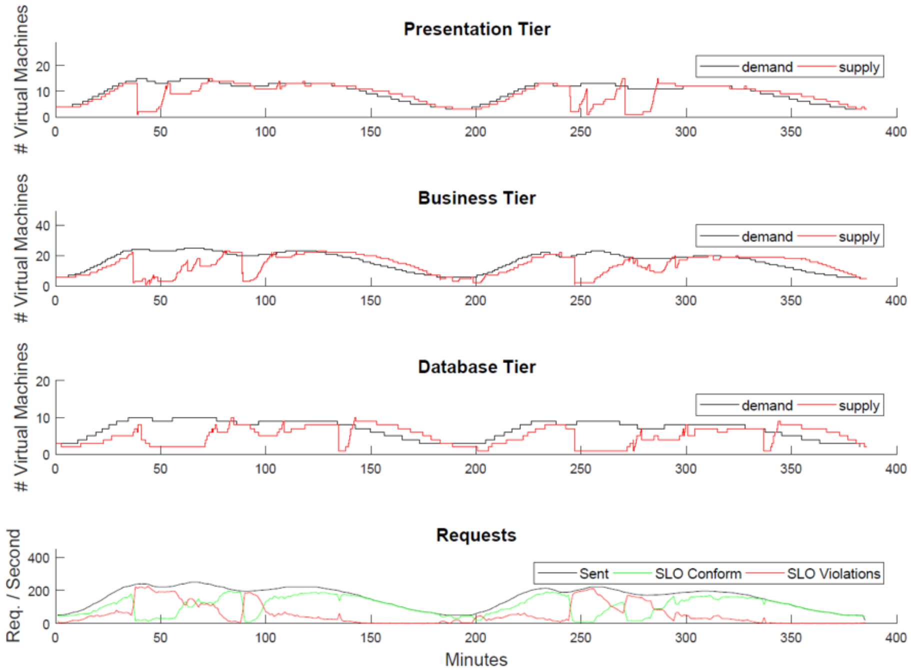


9% SLO violations

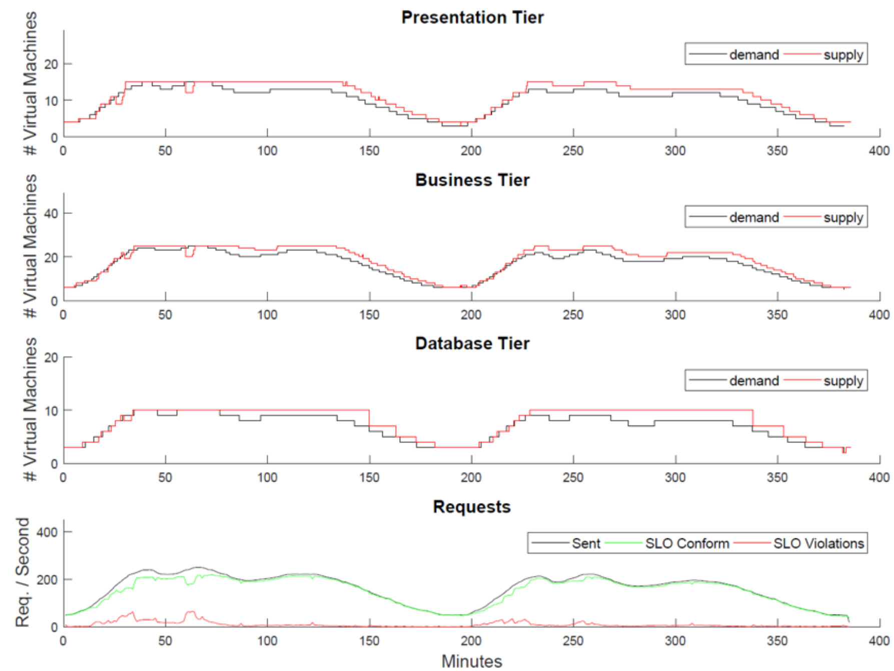
Multi-Tier Auto-Scaling

Model	No Cost	EC2	Diff.
Charged h	263	232	-12%
Accounted h	93	174	+87%
SLO violation	7%	2.6%	-63%

Alternative: Reg



Chameleon



References

- **Bates69**: J. M. Bates and C. W. Granger, “The combination of forecasts,” Journal of the Operational Research Society, vol. 20, no. 4, pp. 451-468, 1969.
- **Clemen89**: R. T. Clemen, “Combining forecasts: A review and annotated bibliography,” International journal of forecasting, vol. 5, no. 4, pp. 559-583, 1989.
- **Menezes00**: L. M. De Menezes, D. W. Bunn, and J. W. Taylor, “Review of guidelines for the use of combined forecasts,” European Journal of Operational Research, vol. 120, no. 1, pp. 190-204, 2000.
- **Collopy92**: F. Collopy and J. S. Armstrong, “Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations,” Management Science, vol. 38, no. 10, pp. 1394-1414, 1992./
- **Wang09**: X. Wang, K. Smith-Miles, and R. Hyndman, “Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series,“ Neurocomputing, vol. 72, no. 10-12, pp. 2581-2594, 2009, lattice Computing and Natural Computing (JCIS 2007) / Neural Networks in Intelligent Systems Designn (ISDA 2007).

References

- **Zhang03**: G. P. Zhang, “Time series forecasting using a hybrid arima and neural network model,” Neurocomputing, vol. 50, pp. 159-175, 2003.
- **Pai05**: P.-F. Pai and C.-S. Lin, “A hybrid arima and support vector machines model in stock price forecasting,” Omega, vol. 33, no. 6, pp. 497-505, 2005.
- **Liu14**: N. Liu, Q. Tang, J. Zhang, W. Fan, and J. Liu, “A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids,” Applied Energy, vol. 129, pp. 336-345, 2014.
- **Züfle17**: M. Züfle, A. Bauer, N. Herbst, V. Curtef, and S. Kounev. “Telescope: A Hybrid Forecast Method for Univariate Time Series,” in Proceedings of the International work-conference on Time Series (ITISE) 2017.

Telescope Takeaway

- Hybrid, multi-step forecasting method for univariate time series

- Benefits:
  - Accuracy improvement for multi-step-ahead forecasting
  - Low variance
  - Reliable time-to-result

- Sweet-Spot:
  - Long, seasonal time series
  - Multiple dominating frequencies
  - Many observations per period

http://descartes.tools/telescope

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

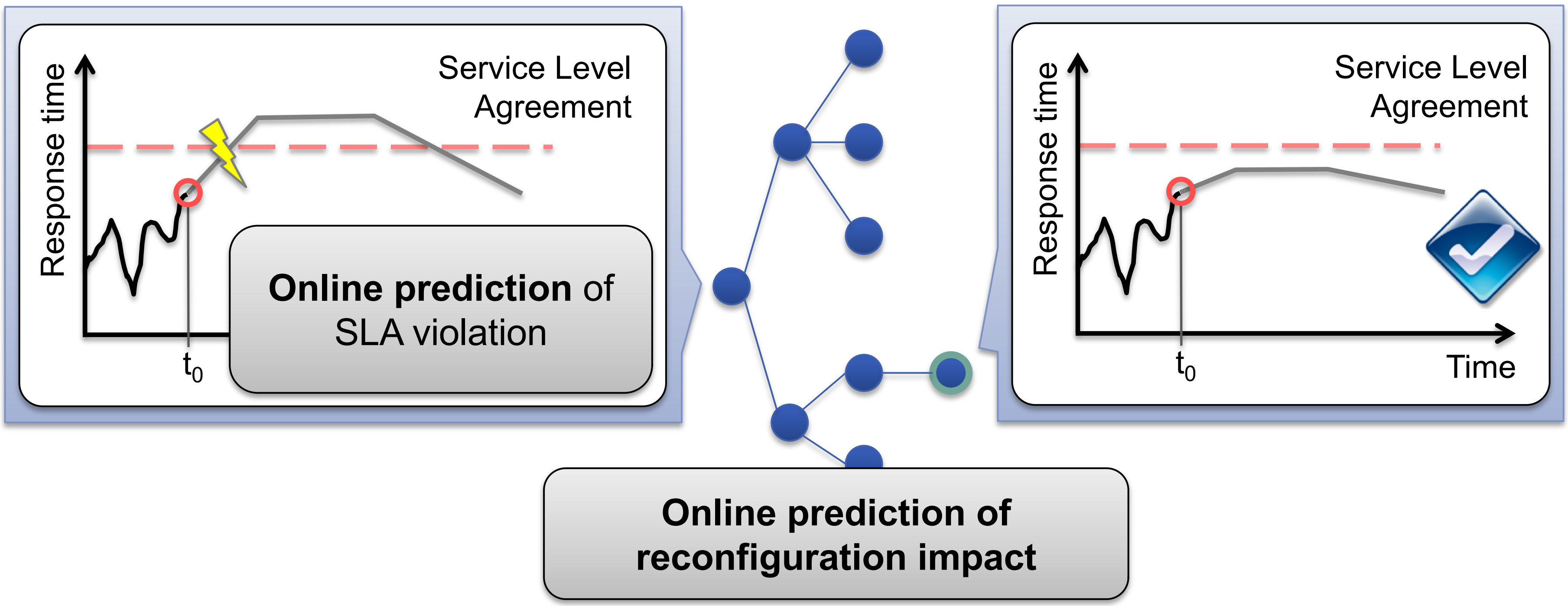
Descartes research

Beyond Forecasting:

Self-Aware Computing in Data Centers

Goal: Self-Aware Data Center

Example Scenario



Descartes Tools

Descartes Modeling Language:

- DML (Descartes Modeling Language)
- DNI (Descartes Network Infrastructures Modeling)

Workload Characterization & Model Extraction:

- LIMBO Load Intensity Modeling Tool
- WCF (Workload Classification and Forecasting Tool)
- LibReDE (Library for Resource Demand Estimation)
- SPA (Storage Performance Analyzer)
- PMX (Performance Model eXtractor)

Declarative Performance Engineering:

- DQL (Descartes Query Language)

Benchmarking:

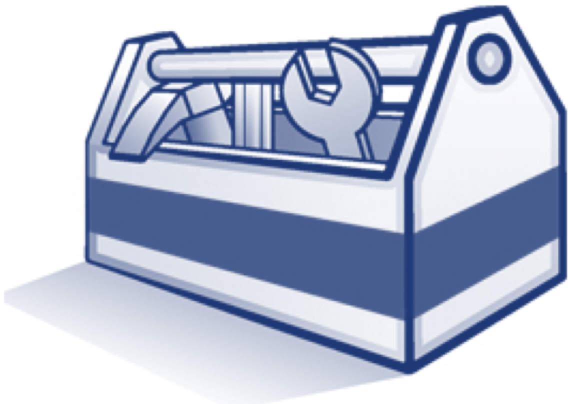
- BUNGEE Cloud Elasticity Benchmark
- hInjector Hypercall Attack Injector

Stochastic Modeling:

- QPME (Queueing Petri net Modeling Environment)

Black-Box Modeling:

- Univariate Interpolation Library



http://descartes.tools

Mailing list available...

Applied Modeling Techniques

Descriptive Architecture-level Models

- OMG Meta Object Facility (MOF)
  - MOF-based meta-models
- (UML MARTE)
- (UML SPT)

Predictive Performance Models

- Bounding techniques
- Operational analysis
- Statistical regression models
- Stochastic process algebras
- (Extended) queueing networks
- Layered queueing networks
- Queueing Petri nets
- Reinforcement learning models
- Detailed simulation models

Workload Forecasting

- AR(I)MA
- Extended exp. smoothing
- tBATS
- Croston’s method
- Cubic smoothing splines
- Neural network-based

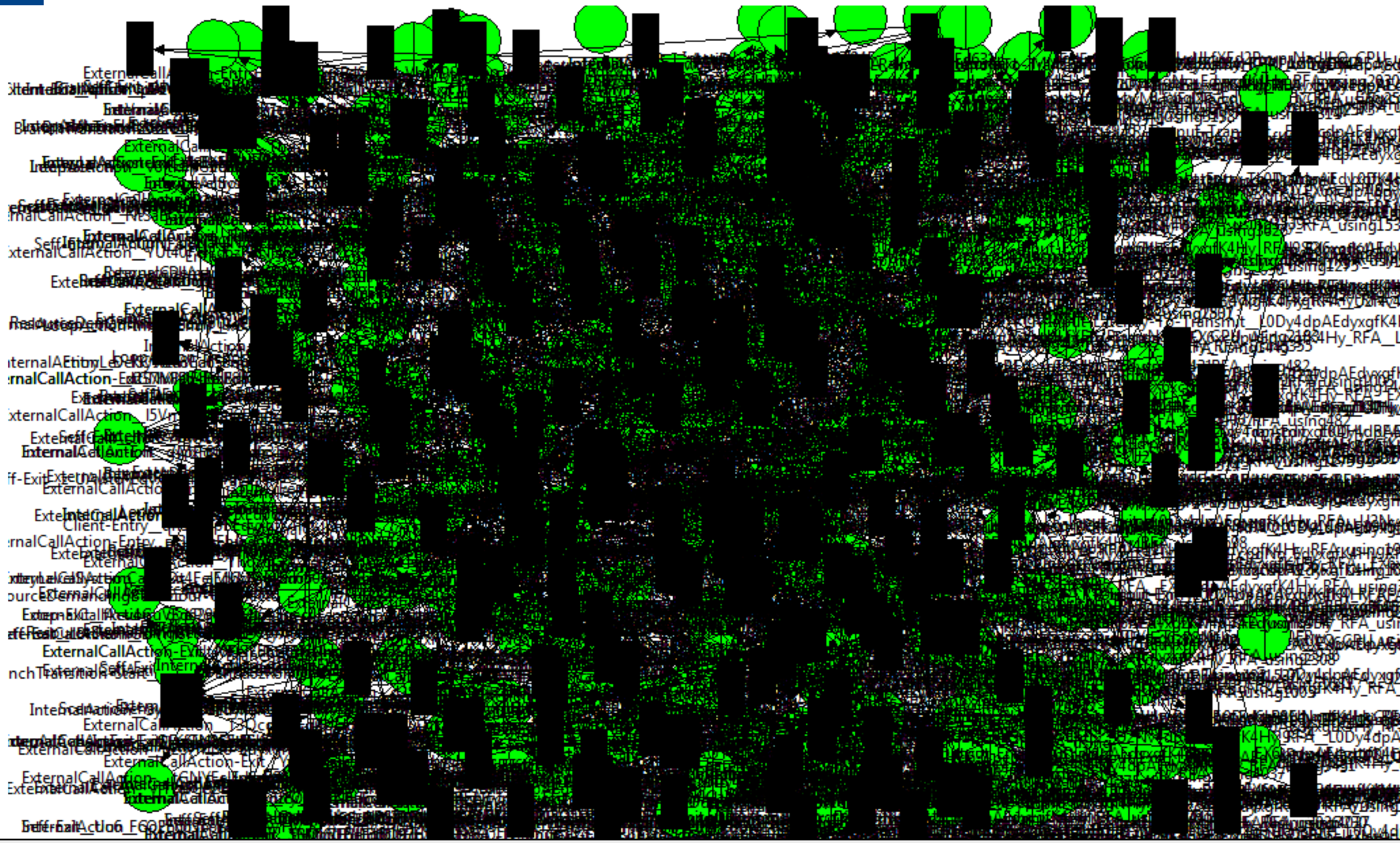
Resource Demand Estimation

- Regression-based techniques
- Kalman filter
- Nonlinear optimization
- Maximum likelihood estimation
- Independent component analysis

Regression Analysis

- MARS
- CART
- M5 trees
- Cubist forests
- Quantile regression forests
- Support vector machines

Case Study: Process Control System (ABB)

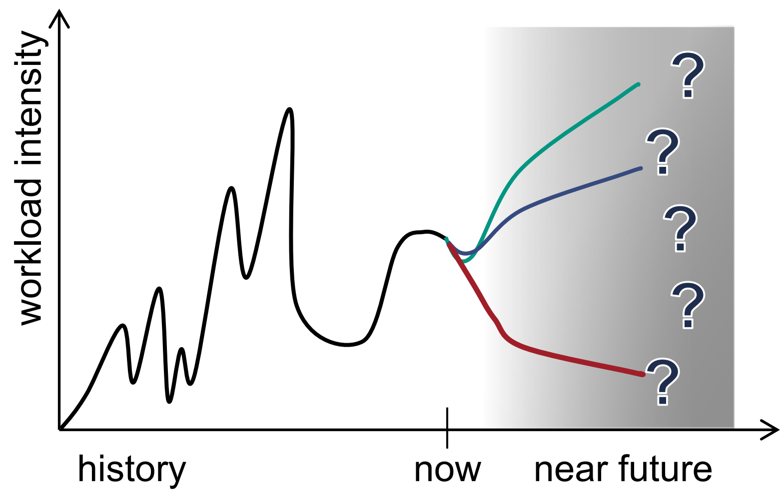


P. Meier, S. Kounev, and H. Koziolek. **Automated transformation of component-based software architecture models to queueing petri nets**. In *19th IEEE/ACM Intl. Symp. on Modeling, Analysis and Simulation of Computer and Telecomm. Systems (MASCOTS), Singapore, July 25-27*, 2011. [.pdf]

LIMBO & TELESCOPE

- Problem:
  - How to capture the load intensity variations (e.g., requests per sec) in a compact mathematical model? → **LIMBO Tool**
  - How to forecast the load intensity (requests per sec) in future time horizons? → **TELESCOPE Tool**

- **Load Intensity Modeling & Forecasting**



http://descartes.tools/limbo
http://descartes.tools/telescope

LibReDE Tool

- Problem: How to estimate the total service time of a given type of request/job at a given resource?

- **Library for Resource Demand Estimation**
  - Ready-to-use implementations of estimation approaches
  - Selection of a suitable approach for a given scenario

http://descartes.tools/librede

S. Spinner, G. Casale, F. Brosig, and S. Kounev. **Evaluating Approaches to Resource Demand Estimation**. *Performance Evaluation*, 92:51 - 71, October 2015, Elsevier B.V. [DOI | http | .pdf]

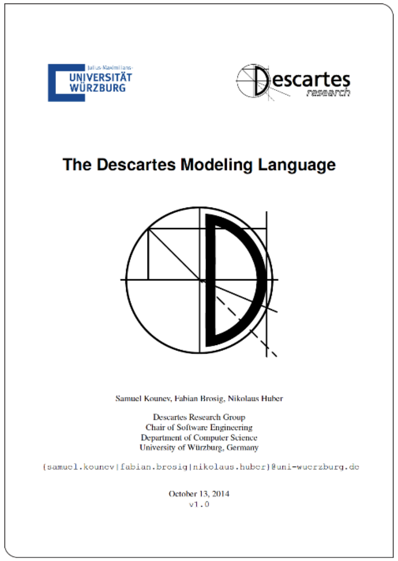
Estimation Approaches

Technique	Variant	References
Approximation with response times		Urgaonkar et al. [13] Nou et al. [14] Brosig et al. [15]
Service Demand Law		Lazowksa [4] Brosig et al. [15]
Linear regression	Least squares	Bard and Shatzoff [16] Rolia et al. [17, 18] Pacifici et al. [19] Kraft et al. [20, 21]
	Least absolute differences	Zhang et al. [22, 23, 24]
	Least trimmed squares	Casale et al. [25, 26]
Kalman filter		Zheng et al. [27, 28] Kumar et al. [29] Wang et al. [30, 31]
Optimization	Non-linear constrained optimization	Zhang et al. [32] Menascé [33]
	Quadratic programming	Liu et al. [34, 35, 36] Kumar et al. [37]
Machine learning	Clusterwise linear regression	Cremonesi et al. [38]
	Independent component analysis	Sharma et al. [39]
	Support vector machine	Kalbasi et al. [40]
	Pattern matching	Cremonesi et al. [41, 42]
Maximum likelihood estimation		Kraft et al. [20] Perez et al. [21]
Gibbs sampling		Sutton and Jordan [43] Wang et al. [44]

S. Spinner, G. Casale, F. Brosig, and S. Kounev. **Evaluating Approaches to Resource Demand Estimation**. *Performance Evaluation*, 92:51 - 71, October 2015, Elsevier B.V. [DOI | http | .pdf]

Descartes Modeling Language

- Problem: How to model performance and resource management related aspects of IT systems and infrastructures at the architecture-level?
  - Prediction of the impact of dynamic changes at run-time
  - Current version focused on performance including capacity, responsiveness and resource efficiency aspects


UNIVERSITÄT WÜRZBURG

Descartes research

The Descartes Modeling Language

Samuel Kounev, Fabian Brosig, Nikolaus Huber

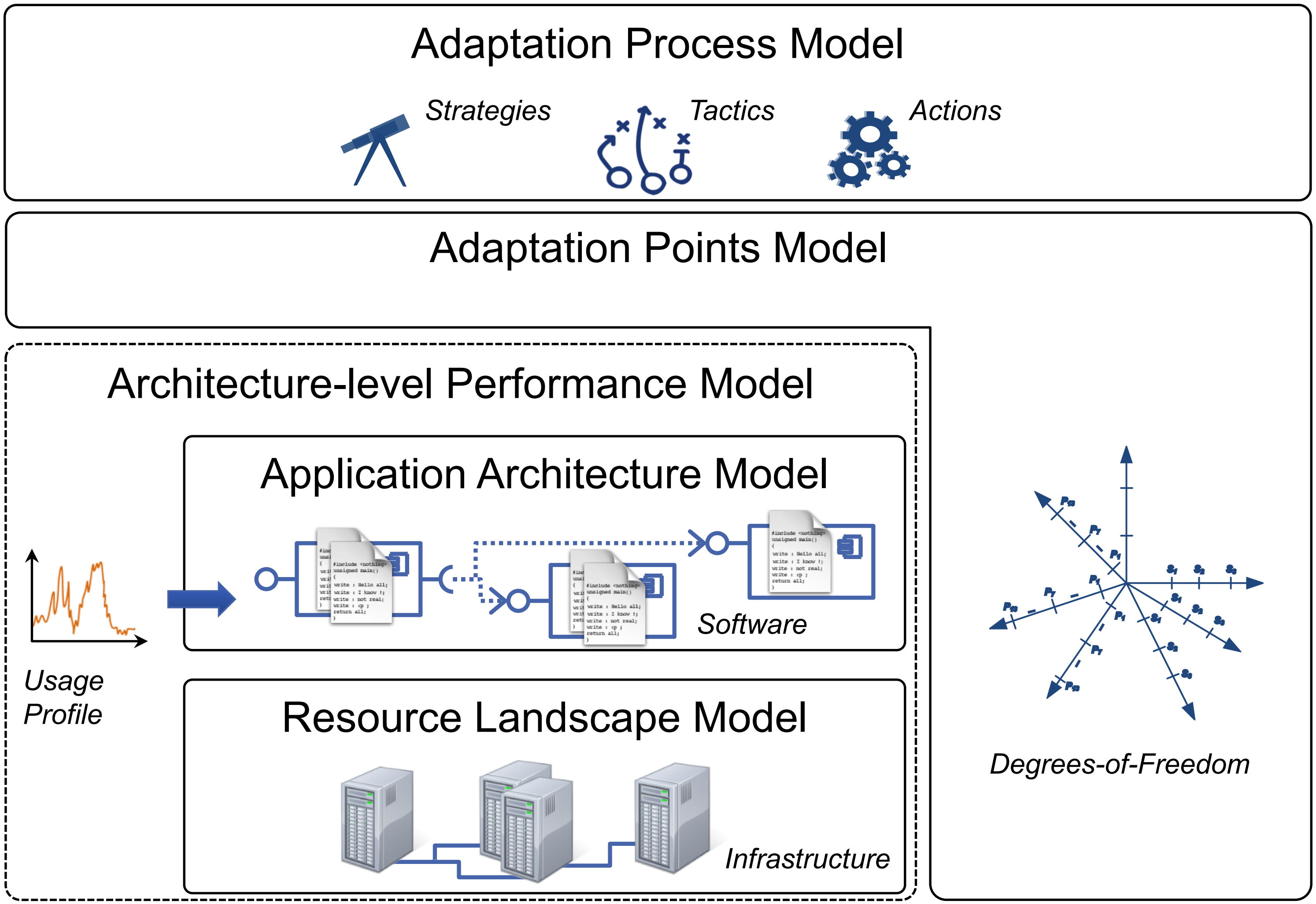
Descartes Research Group
Chair of Software Engineering
Department of Computer Science
University of Würzburg, Germany

{samuel.kounev|fabian.brosig|nikolaus.huber}@uni-wuerzburg.de

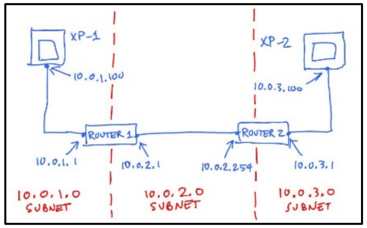
October 13, 2014
v1.0

http://descartes.tools/dml

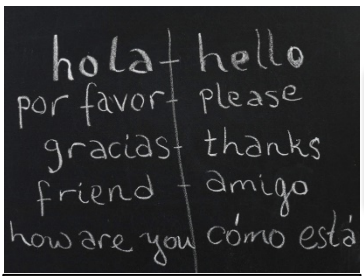
DML Sub-Models



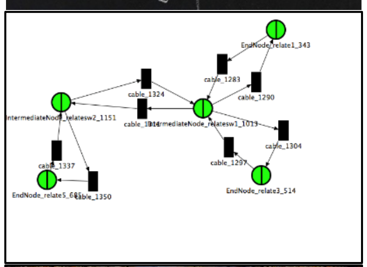
Flexible Modeling of Data Center Networks for Capacity Management



DNI Meta-Model
Generic modeling formalism for SDN- and NFV-based data center networks performance.



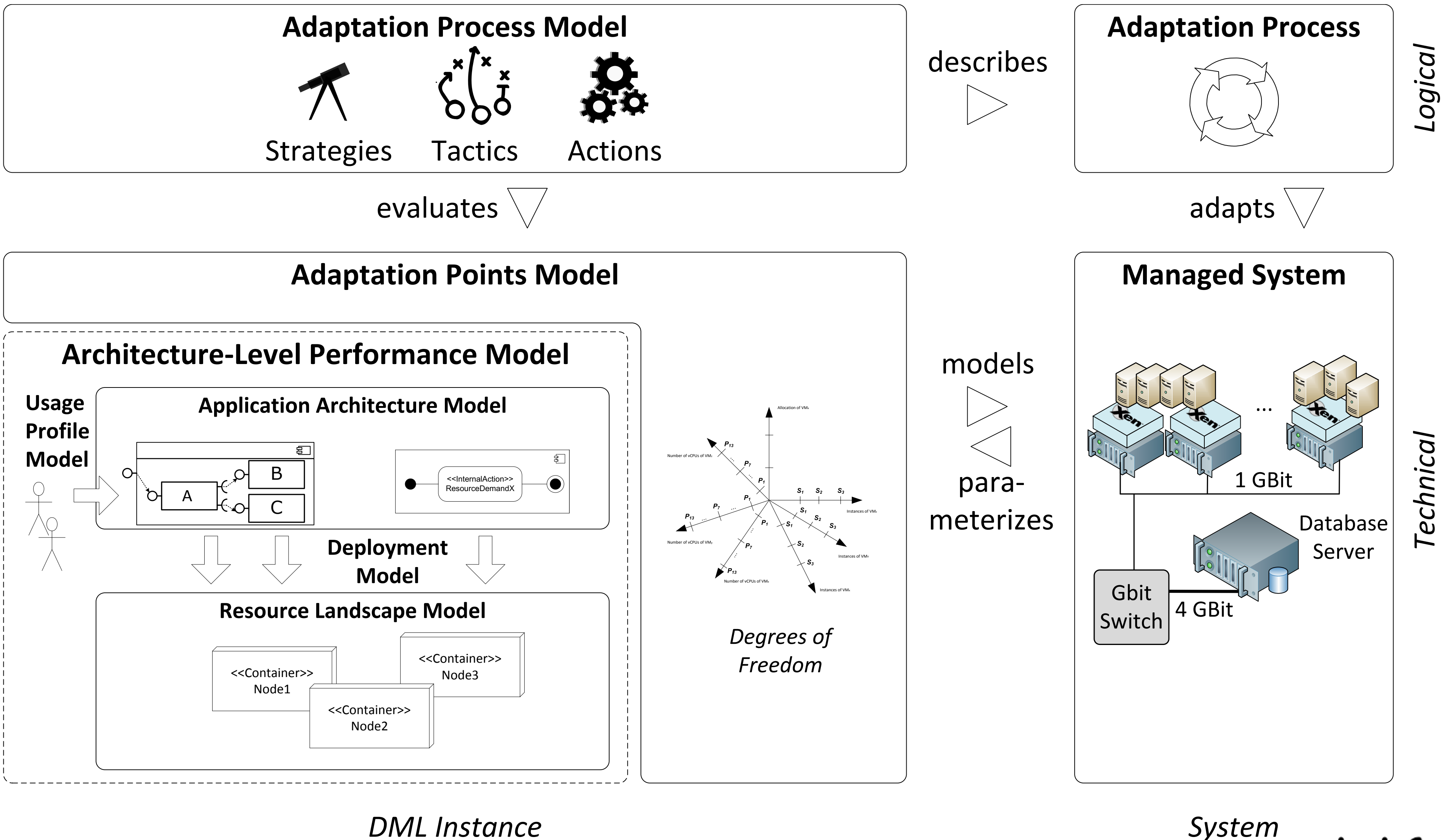
Model Transformations x6
Automated transformations to different predictive models.



Model Solvers ≤10
Solvers supporting trade-offs btw. accuracy and solving time.

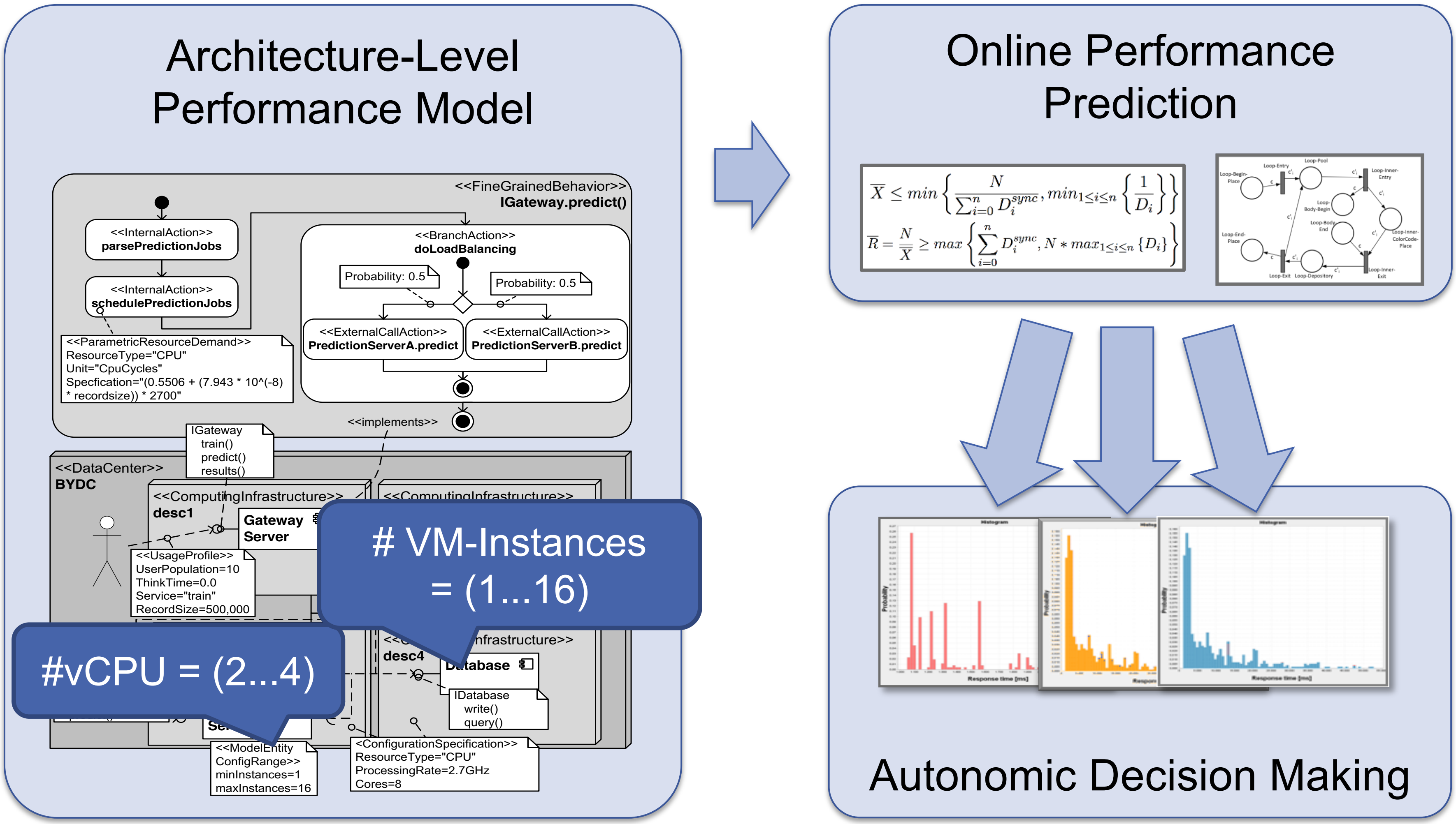
Model Extraction
Traffic models can be extracted automatically from traces.

Big Picture



Adaptation Process Model
Strategies
Tactics
Actions
describes
Adaptation Process
Logical
evaluates
adapts
Adaptation Points Model
Managed System
Architecture-Level Performance Model
Usage Profile Model
Application Architecture Model
B
A
C
<<InternalAction>>
ResourceDemandX
Deployment Model
Resource Landscape Model
<<Container>>
Node1
<<Container>>
Node2
<<Container>>
Node3
Degrees of Freedom
models
para-meterizes
1 GBit
Database Server
Gbit Switch
4 GBit
Technical
DML Instance
System

Online Performance Prediction

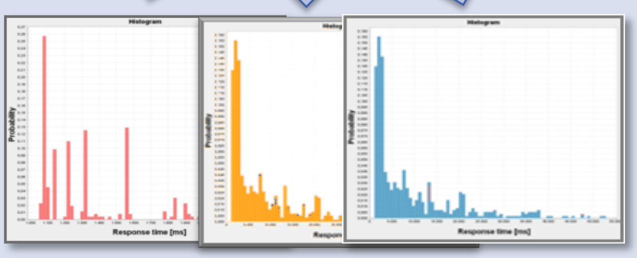


Tailored Model Solution

Analytical Analysis

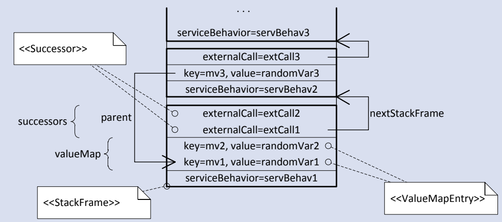
$$R \geq \max\left[N \times \max\{D_i\}, \sum_{i=1}^{K} D_i\right] \qquad X_0 \leq \min\left[\frac{1}{\max\{D_i\}}, \frac{N}{\sum_{i=1}^{K} D_i}\right]$$

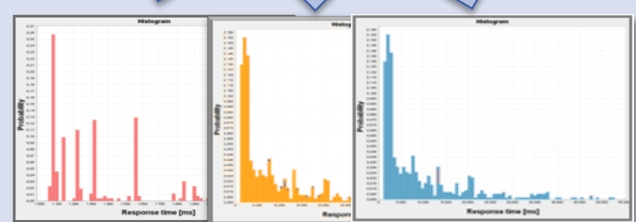
$$\frac{N}{\max\{D_i\}[K+N-1]} \leq X_0 \leq \frac{N}{avg\{D_i\}[K+N-1]}$$



Analysis Results

Simulative Analysis

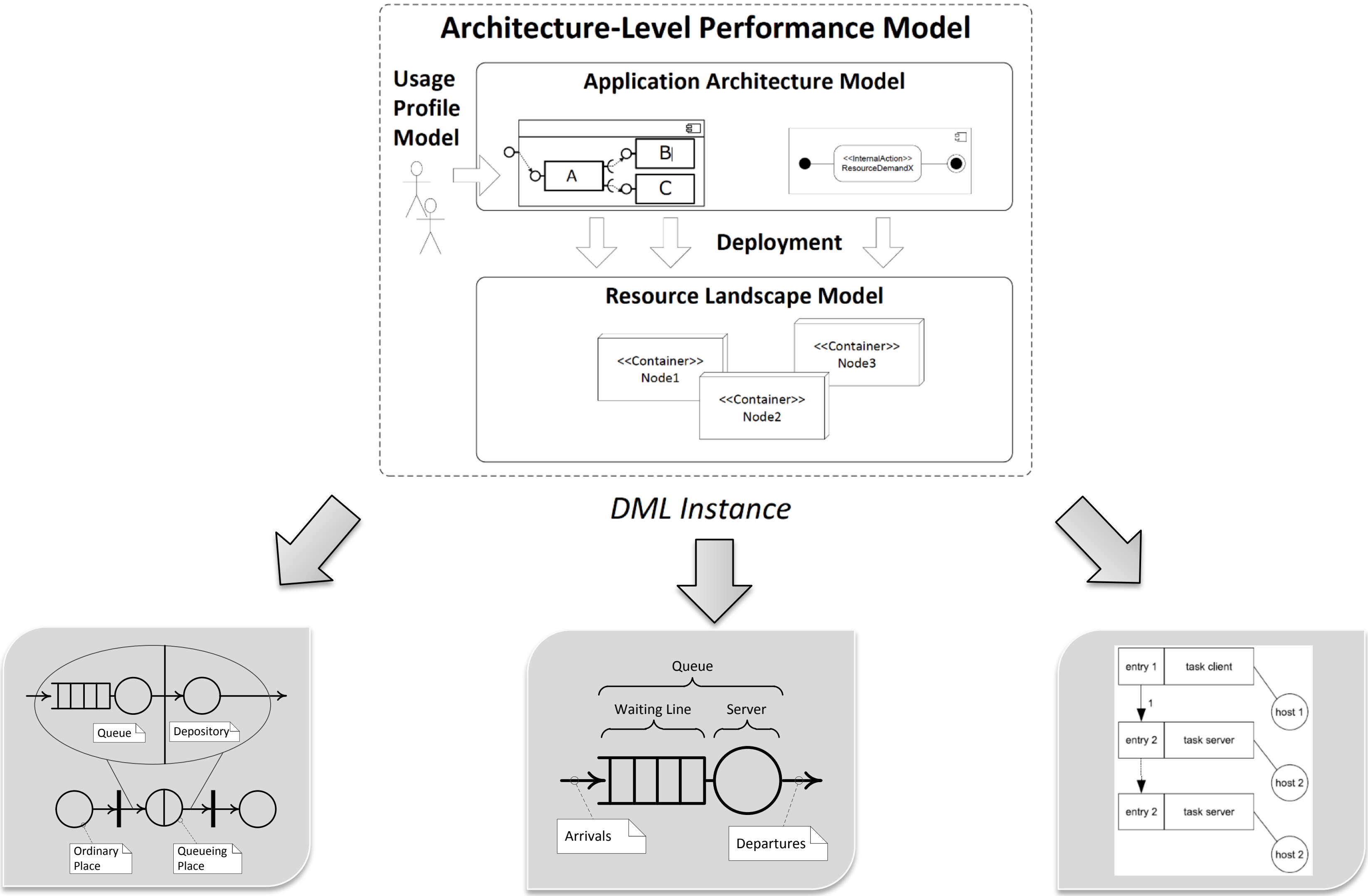




Analysis Results

Fabian Brosig, Philipp Meier, Steffen Becker, Anne Koziolek, Heiko Koziolek, and Samuel Kounev. **Quantitative Evaluation of Model-Driven Performance Analysis and Simulation of Component-based Architectures**. *IEEE Transactions on Software Engineering (TSE)*, 41(2):157-175, February 2015, IEEE. [DOI | http | .pdf]

Transformations to Predictive Models



Queueing Petri Net

Bounds Analysis Model

Layered Queueing Network

Declarative Performance Engineering

```
WHAT IF
    'UserCount' <ADD 100, ADD 200>,
    'CPU.frequency' <MULTIPLY 2.5>,
FILTERING Utilization
USING dml@'specj2010.properties';

SELECT system.bottlenecks
FOR SYSTEM '2ndfs834hkl' AS system
USING dml@'specj2010.properties';


WHAT IF migrateVM(vm1,server2)
FILTERING Utilization
USING dml@'specj2010.properties';
```

J. Walter, A. van Hoorn, H. Koziolek, D. Okanovic, and S. Kounev. **Asking "What?", Automating the "How?": The Vision of Declarative Performance Engineering**. In *Proceedings of the 7th ACM/SPEC International Conference on Performance Engineering (ICPE 2016)*, Delft, the Netherlands, March 12, 2016. [slides | .pdf]

Putting it All Together

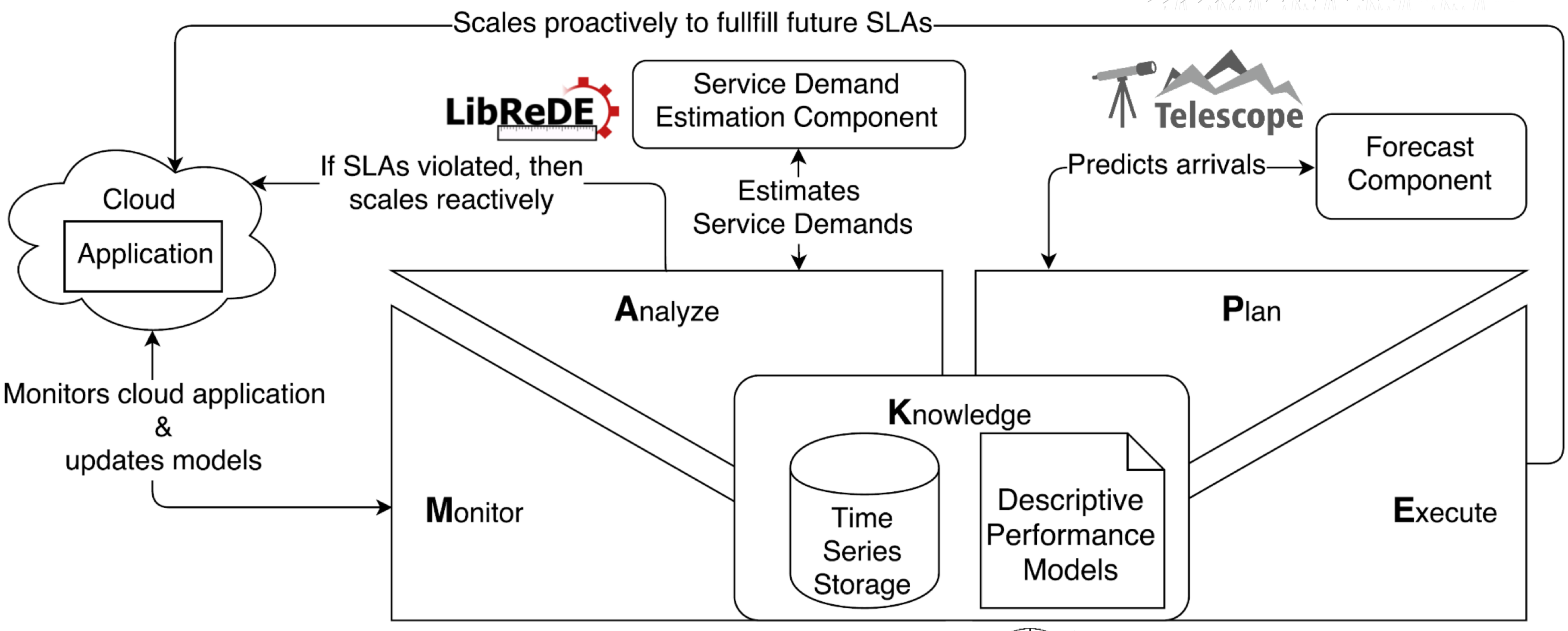
DESIGN AND EVALUATION OF A PROACTIVE, APPLICATION-AWARE AUTO-SCALER

CHAMELEON

http://descartes.tools/chameleon

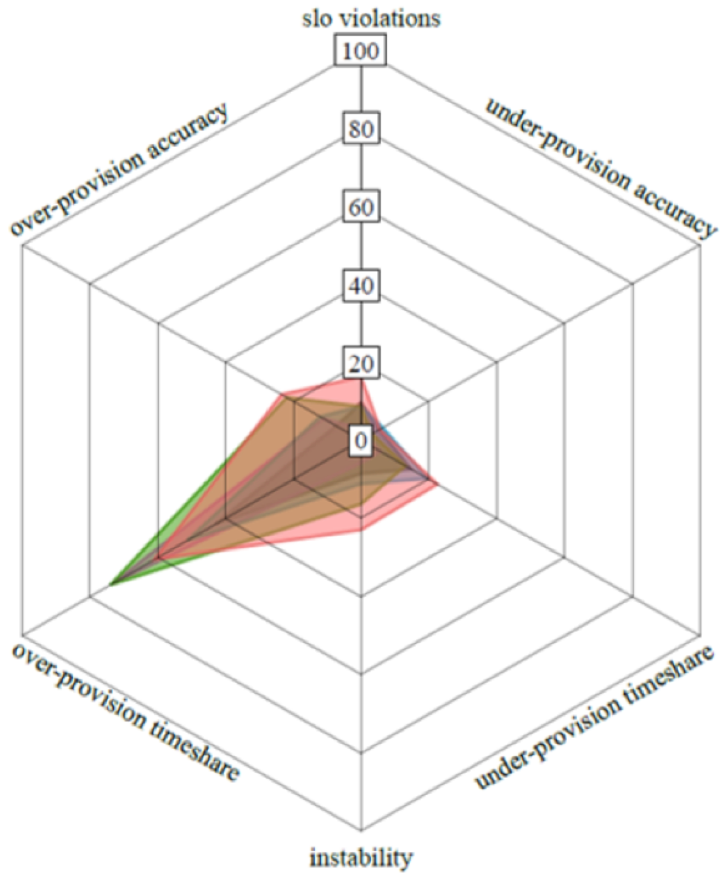
Chameleon

- A Hybrid, Proactive Auto-Scaling Mechanism
  - Reactive and proactive

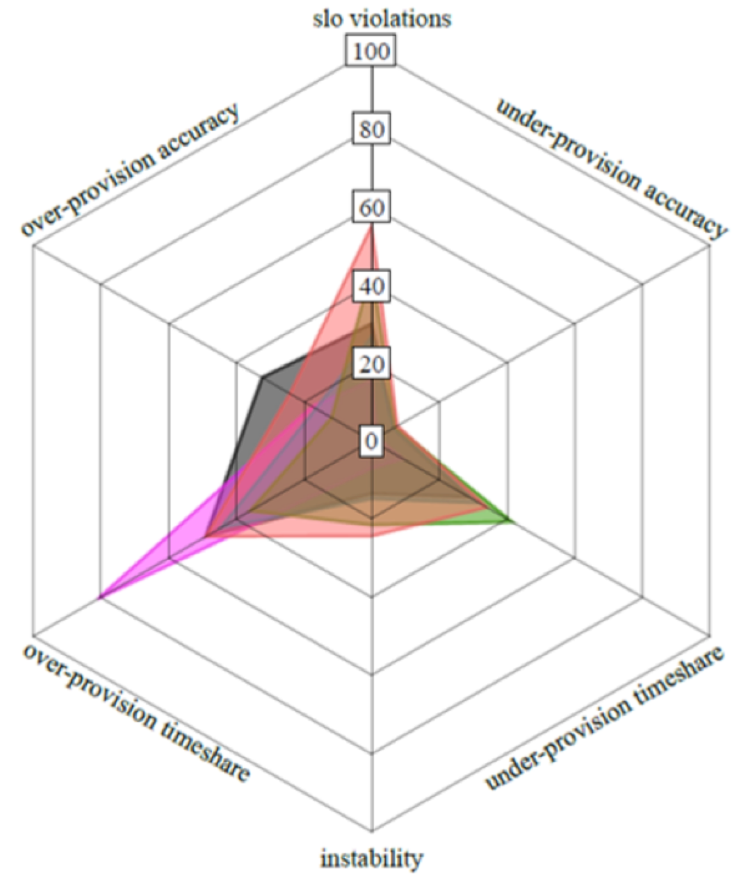


EVALUATION SUMMARY

IBM Transaction
Retailrocket
German Wikipedia
FIFA Worldcup 1998
Bibsonomy



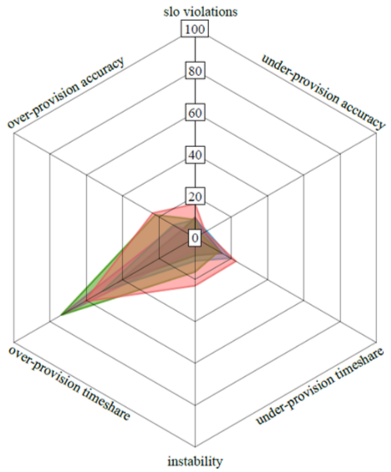
Metric overview Chameleon.



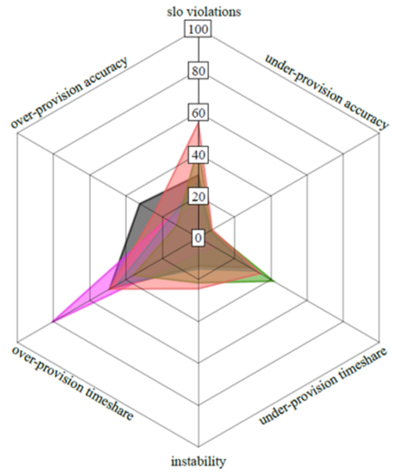
Metric overview Adapt.

The smaller the values (= peaks) the better the performance of the Auto-scaler

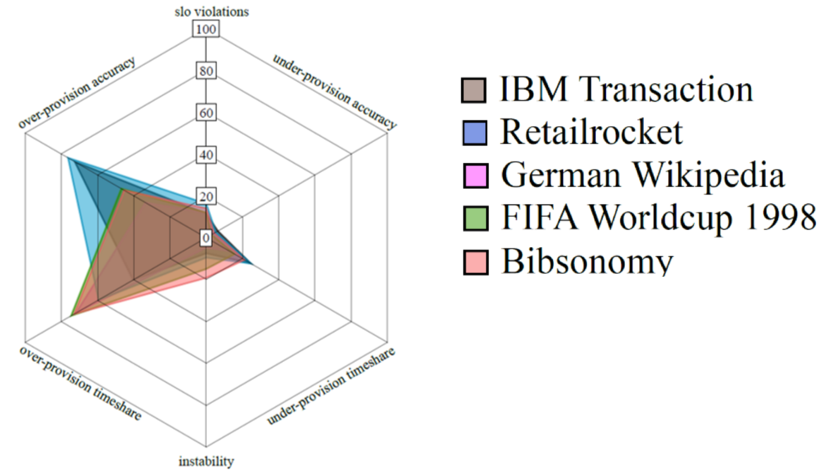
EVALUATION SUMMARY



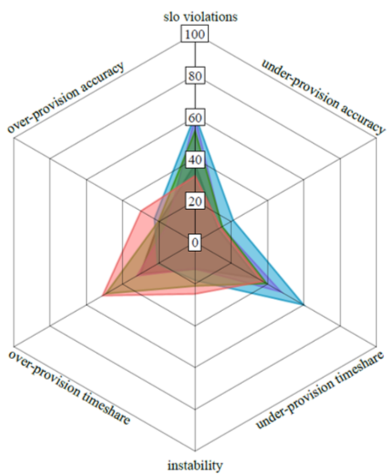
Metric overview Chameleon.



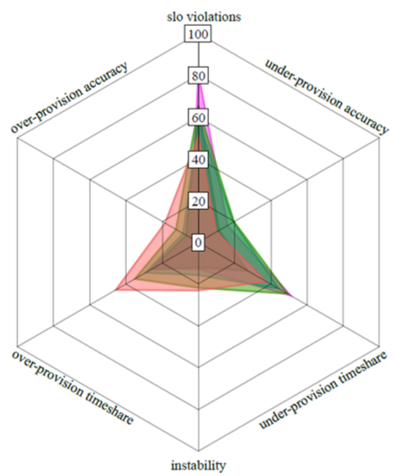
Metric overview Adapt.



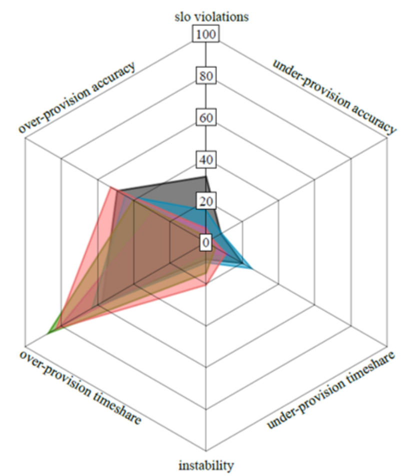
Metric overview Hist.



Metric overview ConPaaS.



Metric overview Reg.



Metric overview Reactive.

Main References

S. Kounev, N. Huber, F. Brosig, and X. Zhu. ***A Model-Based Approach to Designing Self-Aware IT Systems and Infrastructures***. IEEE Computer, 49(7):53–61, July 2016.

N. Huber, F. Brosig, S. Spinner, S. Kounev, and M. Bähr. ***Model-Based Self-Aware Performance and Resource Management Using the Descartes Modeling Language***. IEEE Transactions on Software Engineering (TSE), PP(99), 2017.

IEEE TRANSACTIONS ON SOFTWARE ENGINEERING

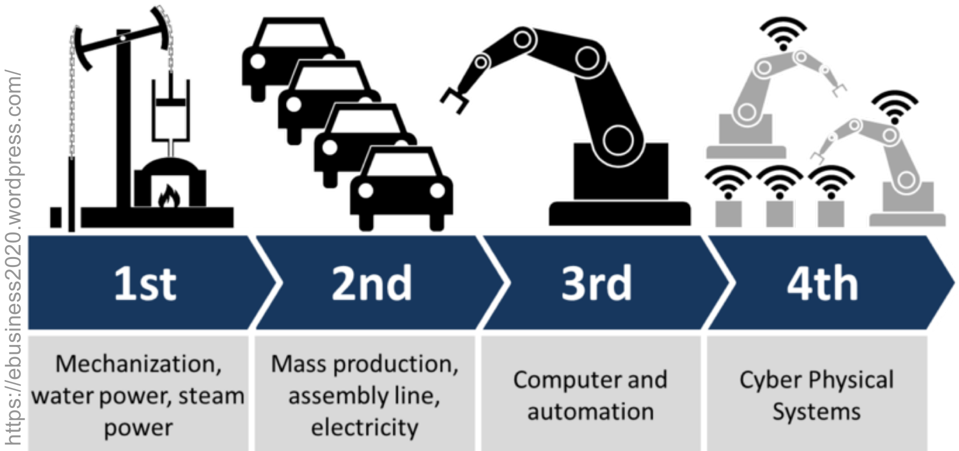
See also ICPE 2017 tutorial at https://se2.informatik.uni-wuerzburg.de/pa/publications/download/slides/1374

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

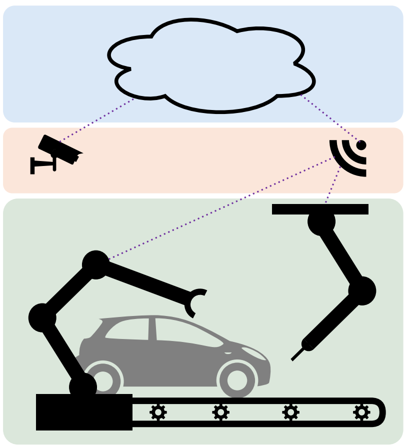
Descartes research

Self-Aware Computing in Industry 4.0

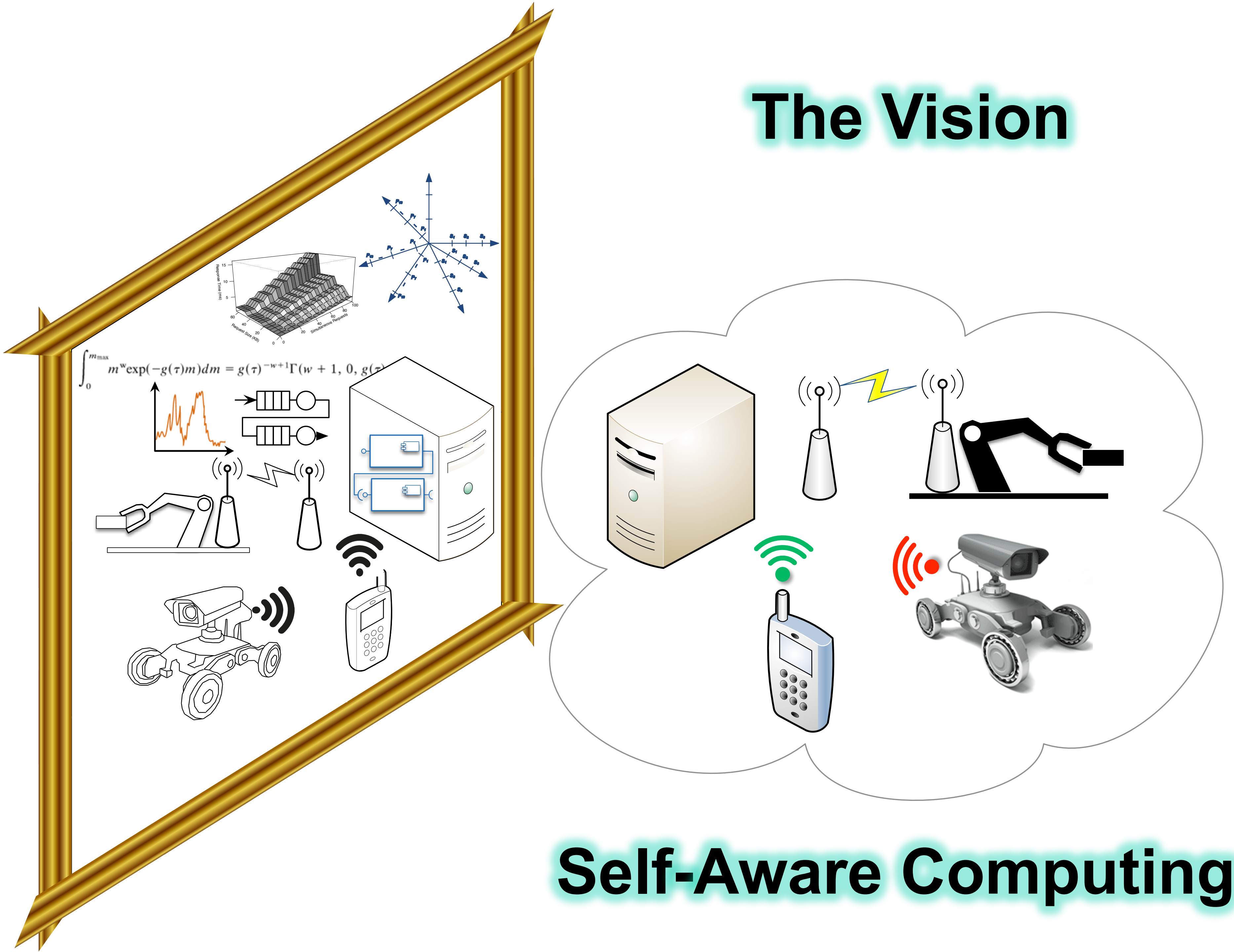
Cyber Physical Systems (CPS)



End of 18th century | Start of 20th century | Start of 1970s | today



"Cyber Physical Systems (CPS) are autonomic systems that fuse the cyber and the physical world. Physical and computing components are connected and communicate by means of a network. The computing components in a CPS process data acquired by monitoring physical processes (e.g. via sensors) and use the result to control them (e.g. via actuators)." [1]


The Vision
Self-Aware Computing

Static Design Knowledge

What is the name, model and version of this machine?

How does it fit in the production workflow?

What production operations does it execute?

What resources, services and other machines does it require to operate correctly?

Knowledge about Current and Past States

What is the machine‘s latest observed workload (jobs/hour)?

What is the machine‘s latest observed performance (job completion time, successful jobs/hour, error rate)?

What is the history of the machine‘s maintenance operations?

What is machine‘s current health status?

Knowledge about Future States

What workload is anticipated / predicted for the next hour / day / week?

What performance would the machine exhibit under this workload?

What is the machine‘s estimated remaining useful life and optimal utilization level?

When should maintenance actions be scheduled?

Sensors in Internet of Things

Das Bild kann nicht angezeigt werden.

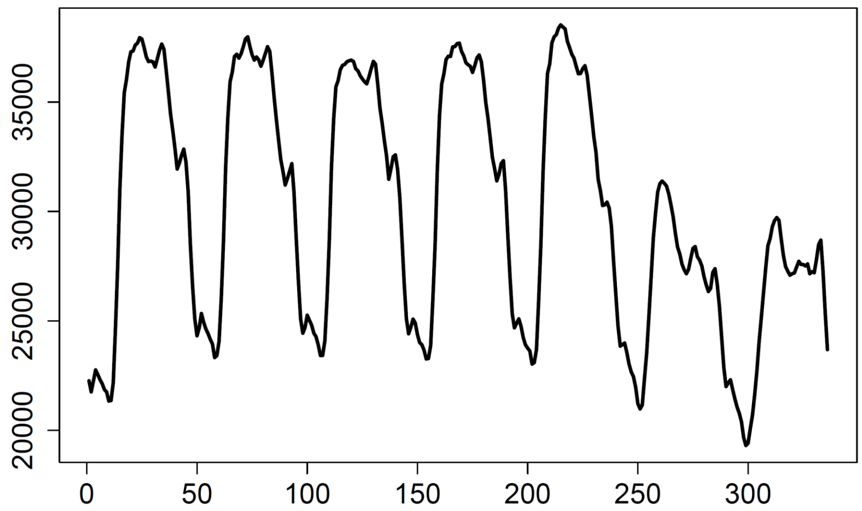
Temperature

Das Bild kann nicht angezeigt werden.

Humidity

Das Bild kann nicht angezeigt werden.

Acceleration



Das Bild kann nicht angezeigt werden.

Das Bild kann nicht angezeigt werden.

Distance

Oscillation

Das Bild kann nicht angezeigt werden.

Light

Das Bild kann nicht angezeigt werden.

Das Bild kann nicht angezeigt werden.

Das Bild kann nicht angezeigt werden.

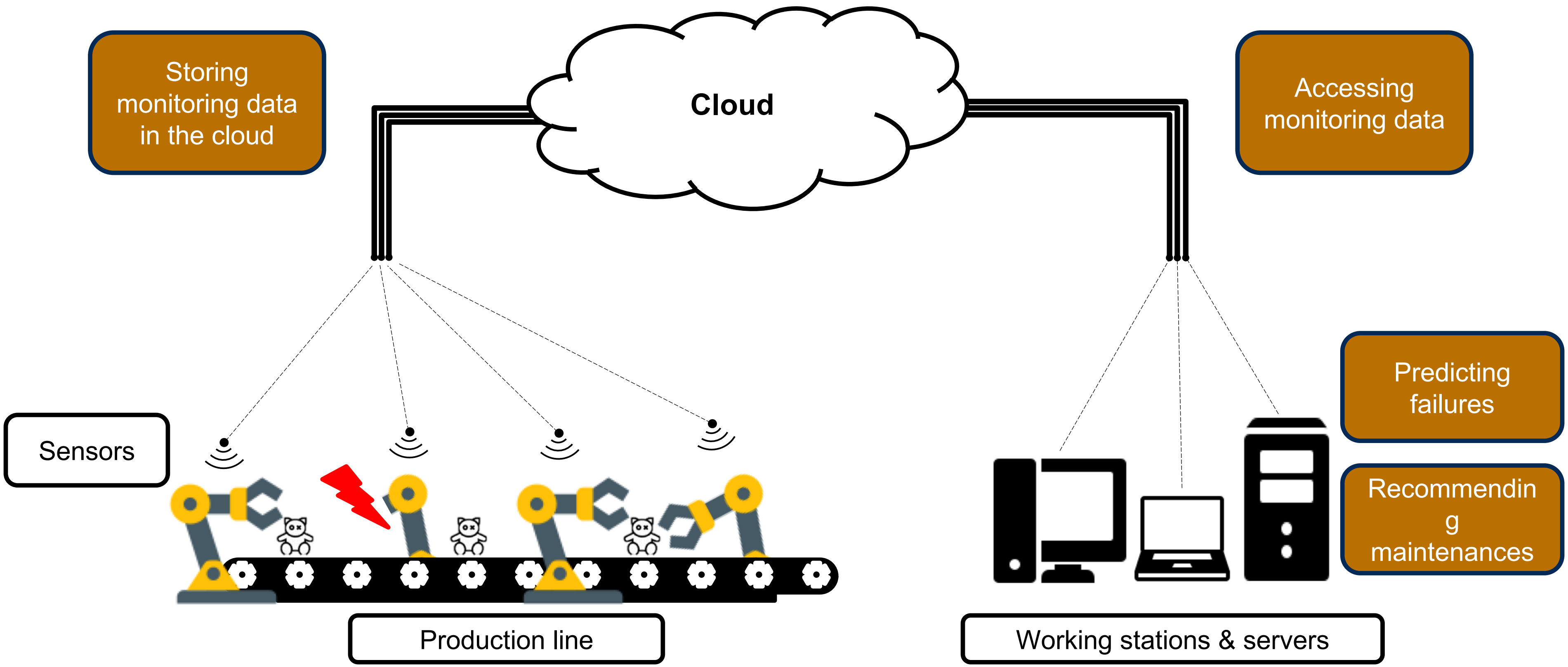
Pressure

Speed

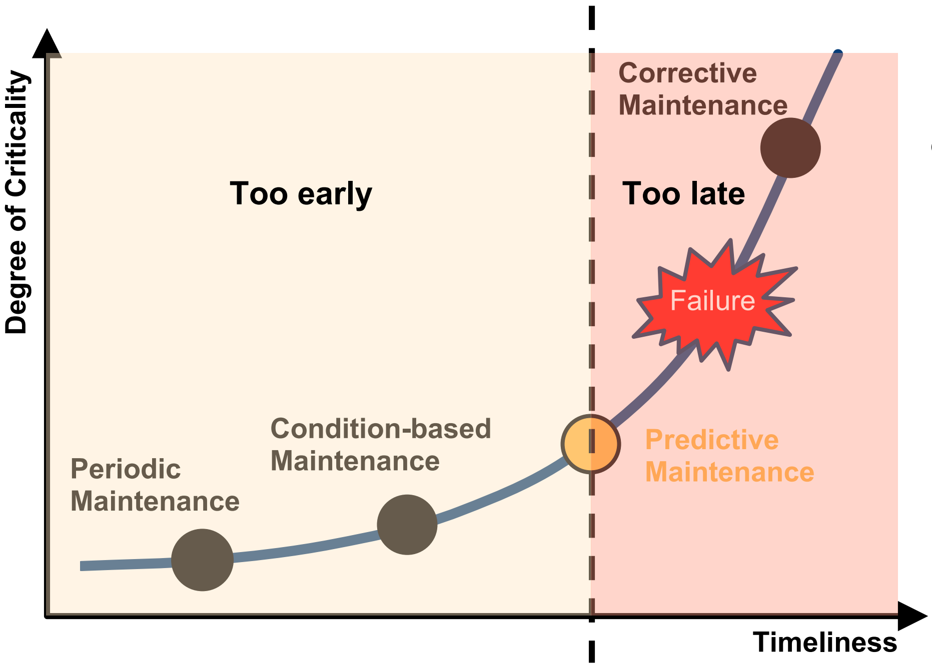
Inclination

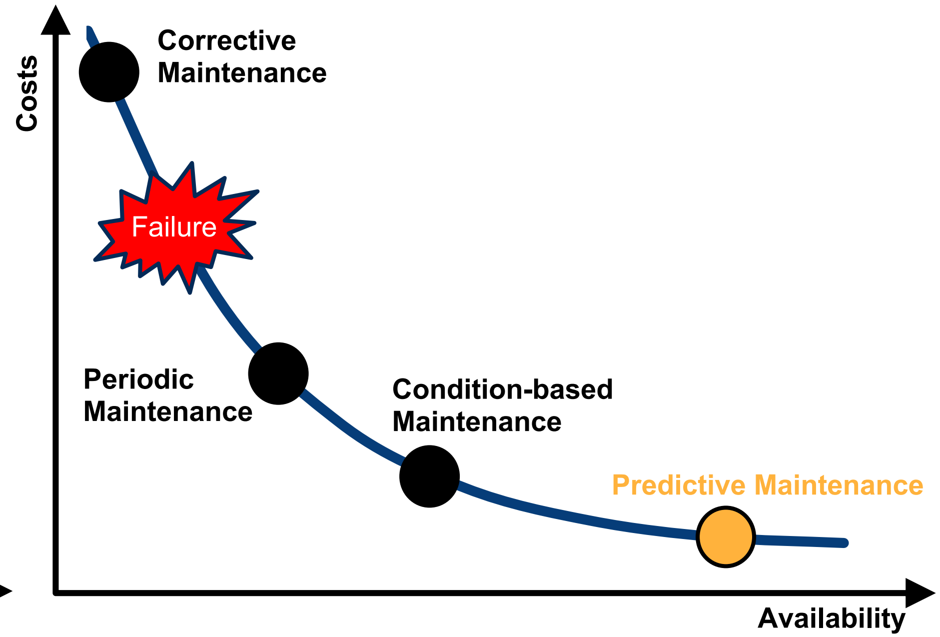
Das Bild kann nicht angezeigt werden.

Predictive Maintenance



Predictive Maintenance

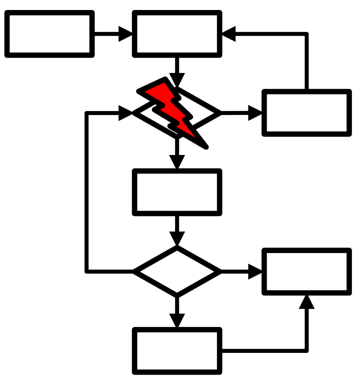


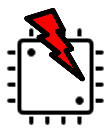


Challenges

Failures as rare events [1]

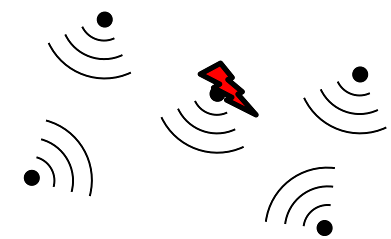
Several types of failures





Lack of real world data with health information

Defect sensors



Huge amount of sensors and data

Inconsistencies

°C

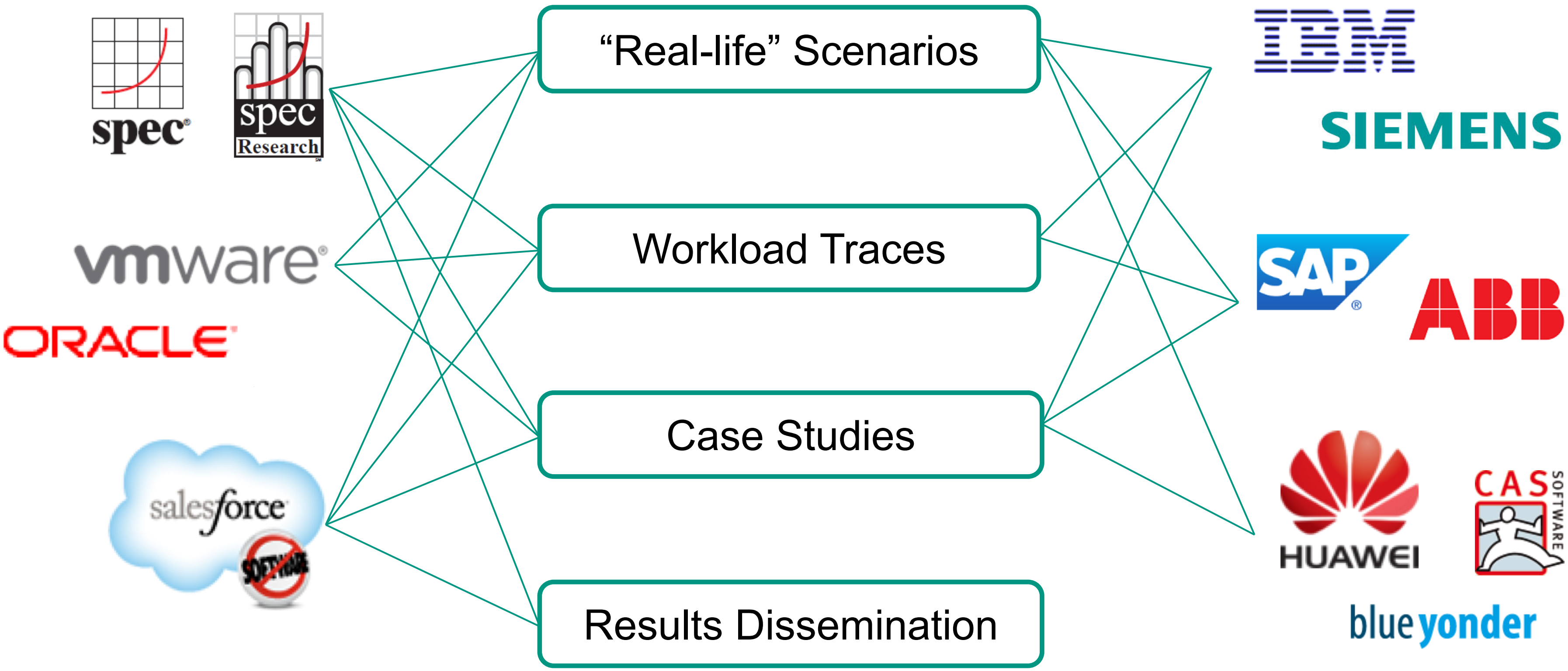
°F

ft in

cm

[1] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." *ACM computing surveys (CSUR)* 41.3 (2009): 15.

Research Partners



Imperial College London
TU Delft
Universidade de Coimbra
Universität Stuttgart
Università della Svizzera italiana
UCD Dublin
Universitat Politècnica de Catalunya
Barcelona Supercomputing Center – Centro Nacional de Supercomputación
Technische Universität Berlin

SPEC Research Group (RG)

- Founded in March 2011: http://research.spec.org
  - Transfer of knowledge btw. academia and industry
- Activities
  - Methods and techniques for experimental system analysis
  - Standard metrics and measurement methodologies
  - Benchmarking and certification
  - Evaluation of academic research results
- 50+ member organizations (selection)

Links for Further Information

- **DML** – Descartes Modeling Language (homepage, publications)
- **DML Bench** (homepage, publications)
- **DQL** – Declarative query language (homepage, publications)
- **DNI** – Descartes network infrastructure modeling (homepage, publications)
- **LibReDE** - Library for resource demand estimation (homepage, publications)
- **LIMBO** – Load intensity modeling tool (homepage, publications)
- **WCF** – Workload classification & forecasting tool (homepage, publications)
- **BUNGEE** – Elasticity benchmarking framework (homepage, publications)
- **hInjector** – Security benchmarking tool (homepage, publications)
- **Further relevant research**
  - **http://descartes-research.net/research/research_areas/**
  - **Self Aware Computing** (publications)

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Descartes
research

Questions?

skounev@acm.org

http://descartes.tools

http://descartes-research.net